AD_____

AWARD NUMBER DAMD17-97-1-7286

TITLE: Detection of Mutations Using a Novel Endonuclease

PRINCIPAL INVESTIGATOR: Anthony T. Yeung, Ph.D.

CONTRACTING ORGANIZATION: Fox Chase Cancer Center
Philadelphia, Pennsylvania 19111

REPORT DATE: July 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Material Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1999	Final (1 Jul 97 - 30 Jun 99)

4. TITLE AND SUBTITLE
Detection of Mutations Using a Novel Endonuclease

5. FUNDING NUMBERS
DAMD17-97-1-7286

6. AUTHOR(S)
Anthony T. Yeung, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fox Chase Cancer Center
Philadelphia, Pennsylvania 19111

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We discovered a nuclease, CEL I from celery, that has high specificity for DNA mismatch, including base-substitutions, insertions and deletions. It has been used to develop a robust mutation detection assay. The substrate is fluorescently labeled PCR product heteroduplexes in which the two strands are of two different colors. CEL I nicks one strand of some molecules to produce a truncated DNA band visualized by an automated DNA sequencer fluorescence-based Genescan analysis. In other molecules, the enzyme nicks at the mismatch in the other strand, and produces a band of the second color. The sizes of the two DNA bands independently pinpoint the location of the mismatch. We have optimized this assay, and demonstrated its utility in the screening of mutations and polymorphisms. Our test samples include all the coding exons of the *BRCA1* gene of 10 patients, 500 basepairs of exon 11 of 100 patients, and all the exons of the *ARSC* gene of 100 individuals. Both of these genes are important to breast cancer. To further economize the assay, we validated a protocol to use a single pair of fluorescent primers in a nested PCR approach to screen all the exons of the ARSC gene of 100 individuals.

14. SUBJECT TERMS
Breast Cancer Mutation detection, BRCA1, Genescan

15. NUMBER OF PAGES
177

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18 298-102

USAPPC V1.00

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

TY Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

TY In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

TY In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Anthony Young 8/9/99

PI - Signature Date

TABLE OF CONTENTS

1. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research, 26*, 4597-4602.
2. Oleykowski, C. A., Bronson Mullins, C. R., Chang, D. W., and Yeung, A. T. (1999) Incision at nucleotide insertions/deletions and basepair mismatches by the SP Nuclease of Spinach. *Biochemistry* **38**, 2200-2205.
3. Yang, B., Wen, X., Oleykowski, C. A., Miller, C. G., Kulinski, J. A., Besack, D., and Yeung, A. T. (1999) Purification and Characterization of the CEL I Endonuclease that has high specificity for mismatch. Submitted to *J. Biol. Chem.* for publication.
4. Kodali, N.S., Oleykowski, C.A., Kowalski, D., Yang, B., Miller, C.G., Besack, D., Kulinski, J.A., and Yeung, A.T. (1999) A Comparison of the CEL I Endonuclease with the Mung Bean Nuclease, two nucleases of the S1 superfamily. Submitted to *J. Biol. Chem.* for publication.
5. Kulinski, J. A., Besack, D., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation, draft enclosed.
6. Besack, D., Kulinski, J. A., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) Polymorphisms of the human steroid sulfatase gene. Manuscript in preparation, draft enclosed.
7. U.S. patent, number 5869245 (1999) "Mismatch endonuclease and its use in identifying mutations in targeted polynucleotide strands".

INTRODUCTION:

The detection of mutations in genes that predispose a person to breast cancer is important towards achieving the prevention and early treatment of breast cancer. My laboratory has discovered a new endonuclease, CEL I from celery, that is highly specific for basepair mismatches and loops in the DNA helix (1). Using CEL I, we have developed a method for mutation detection that may replace SSCP (single-strand conformation polymorphism) and many other currently used mutation detection methods (2-10). In this CEL I mutation detection approach, fluorescently labeled heteroduplexes of DNA are prepared from PCR (polymerase chain reaction) products of different alleles of the *BRCA1* gene of research participants (11). When CEL I cuts these heteroduplexes at the sites of mismatches, new DNA bands are observed in automated DNA GeneScan analysis. This method can detect mutations effectively, including those that are in close proximity to DNA polymorphisms, and in DNA targets as large as 3,000 basepairs. The system may be of great value to the prediction and early detection of cancer. This report presents the optimization of this assay and its application to a number of studies.

BODY:

The purpose of the proposed research is to further develop and document the efficacy of the CEL I mutation detection assay. The rational is that by in depth understanding of the CEL I endonuclease, the CEL I mutation detection assay can be optimized. Secondly, by testing the CEL I mutation detection assay on samples from a large number of people, we can find out the efficacy of the assay. Above plan was parsed into the Statement of Work originally envisioned for this proposal. As work progressed, it became obvious that some of the planned approaches should not be pursued the way they were stated. Based on new experience, some Statements of Work have become lower in priority because of new findings. The original proposal reviewers were also correct in their comments that the stated goals were too extensive to be completed with the limited manpower and budget allowed by this grant. In spite of above limitations, we have made substantial progress in most of the stated tasks and in other tasks that we deem important to the spirit of the proposal. To make that possible, we have enlisted manpower support from funds provided by another source. Important tools made available to us include an ample supply of CEL I from the purification of CEL I to homogeneity from 100 Kg of celery, and the availability of ample patient DNA samples from high risk family members through the Margaret Dyson/Family Risk Assessment Program (FRAP). Individuals participating in FRAP have agreed to allow their samples to be used for a wide range of research purposes including screening for mutations in candidate cancer predisposing genes, such as *BRCA1*. The participating individuals had previously been screened for *BRCA1* mutations by the Clinical Genetic Testing laboratory at FCCC, the results of which were later confirmed by sequencing. However, CEL I mutation detection in our current study was done in a blind fashion. Some of the testing of CEL I mutation detection method have also been shifted from our lab to numerous user labs internationally. In fact, one user has developed the CEL I assay to allow the detection

of a mutation located within a 70 Kbp DNA region in a single assay (personal communication). Specific changes in our plan in order to achieve the stated goals of the proposal are detailed in the report below:

The report will follow the form of the original Statement of Work:

Specific Aim 1. Evaluate CEL I mutation detection protocol with 120 research participant samples. Evaluate potential limitations in CEL I mutation detection: sequence context, fragment size, and PCR quality

Task 1: Months 1-6 Screening of first 30 research participants using current mutation screening procedure while new methods are being developed.

This part of the study was done in collaboration with Dr. Andrew Godwin of the Clinical Genetic Testing laboratory of our institution, and performed on over 200 patients. The assay was not exclusively using CEL I nuclease. T4 endonuclease VII system was also done in parallel with most samples, and the results were confirmed with DNA sequencing. This data is thus far confidential and cannot be provided here is detail.

Task 2: Months 7-12 Screening second 30 research participants using the newer protocols developed in months 1-6.

This task was performed as for Task 1. Optimization of the assay is described in the following manuscript:

Kulinski, J. A., Besack, D. , Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation, draft is enclosed.

Briefly, CEL I was found stable to freeze thawing for 20 cycles, and for long term storage. The assay was shown to be robust, tolerant of a wide spectrum of salts, buffers, enzyme concentrations, substrate concentrations, and incubation times.

Task 3: Months 13-18 Screening third 30 research participants with newer protocols while testing newer protocols including fluorescence energy transfer primers.

This task was performed as for Task 1 except with the newer protocols established and described in this report. Fluorescence energy transfer primers, in particular, were not used as explained below.

Task 4: Months 19-24 Screening fourth 30 research participants with final protocols optimized in this study.

This task with the final optimized protocol is reported in detail as the second part of the enclosed manuscript:

Kulinski, J. A., Besack, D., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation, draft is enclosed.

Two studies are reported to demonstrate the utility of the assay: (i) the new streamlined protocol was used to rapidly evaluate all the exons of the *BRCA1* gene of 10 persons for mutations and polymorphisms, (ii) A *BRCA1* exon of a 500 bp region of 100 people was evaluated for mutations and polymorphisms in a single DNA sequencing gel. The assay used multiplexing of the DNA of five people in each mutation detection reaction for one DNA sequencing lane.

The study is data intensive. Testing of the 10 persons involved over 10 Genescan gel runs, over 300 PCR amplifications experiments and 300 CEL I digestions. The following pages is a presentation of the data set of the mutation detection of the complete coding region of the *BRCA1* gene of one individual, and the data set of mutations and polymorphisms detected in the other nine individuals. This presentation is not included in the manuscript due to space restrain. In these figures, B=background peaks from PCR artifacts. Mutations and polymorphisms are indicated with the nucleotide changes denoted above the peaks. Note the high frequency of multiple polymorphisms and mutations appearing within a single person in a region as small as 600 bp of exon 11 of the *BRCA1* gene. The ability to reveal these multiple mismatches within a small DNA fragment is a strength of the CEL I mutation detection system.

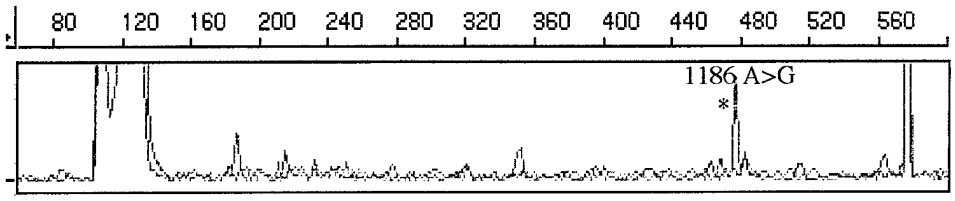
Lists of polymorphisms found in 10 individuals

Polymorphism	Individual 1	2	3	4	5	6	7	8	9	10
coding										
1 [exon 11] 1186 A>G	•		•	•	•	•			•	•
2 [exon 11] 2196 G>A		•						•	•	
3 [exon 11] 2201 C>T				•	•	•		•	•	
4 [exon 11] 2430 C>T				•	•	•			•	
5 [exon 11] 2731 C>T				•	•	•		•	•	
6 [exon 11] 3233 A>G						•			•	
7 [exon 11] 3667 A>G				•	•	•		•	•	
8 [exon 13] 4427 C>T				•	•	•		•	•	
9 [exon 16] 4956 A>G				•	•	Δ		•	•	
intronic										
10 [exon 17 -92] A>G				•	•	•		•	•	
11 [exon 18 +66] G>A				•	•	•		•	•	

List of polymorphisms found in the screening of the human BRCA1 gene of 10 individuals. The first column lists the nucleotide number and base change of the polymorphism. For intronic base changes, the exon closest to the base change is given with the number of bases (+ or -) into the intron the change occurs. A total of 50 (11 different) polymorphisms were found.

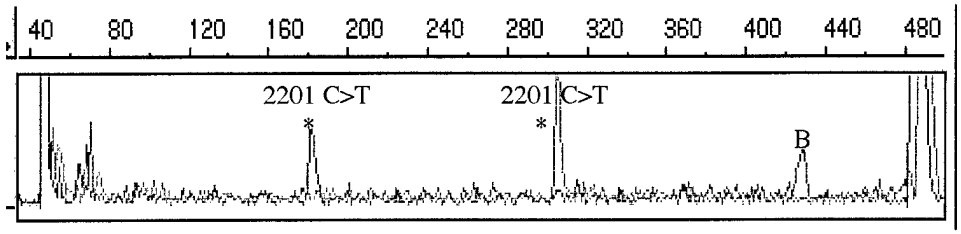
Δ Data not conclusive due to sample error. There was no DNA in the lane and sample will be re-done.

POLYMORPHISM DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #1



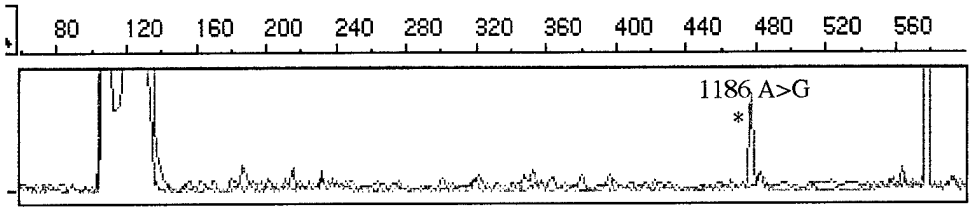
Exon 11
Sect. 2

POLYMORPHISM DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #2



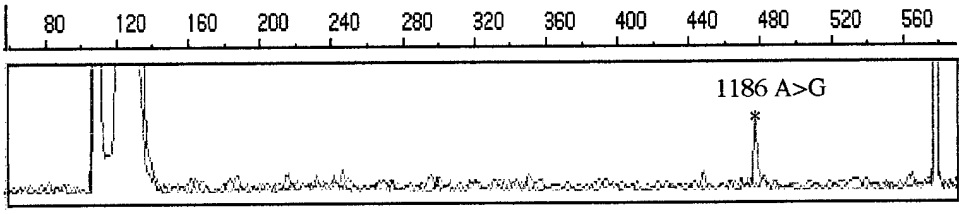
Exon 11
Sect. 4

POLYMORPHISM DETECTED BY CEL I MUTATION DETECTION
FOR INDIVIDUAL #3

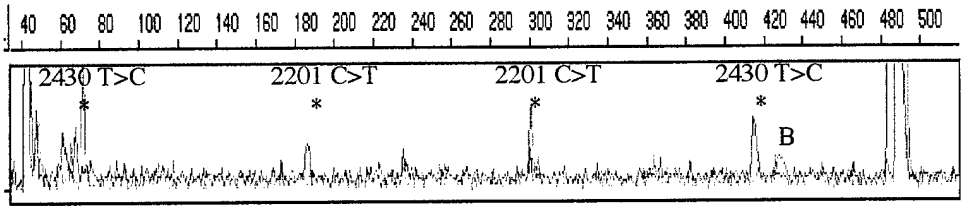


Exon 11
Sect. 2

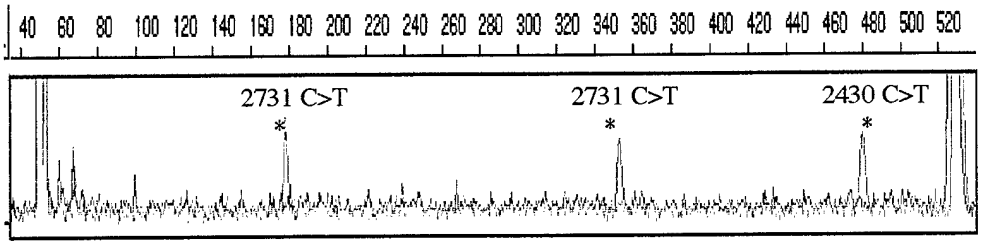
POLYMORPHISMS DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #4



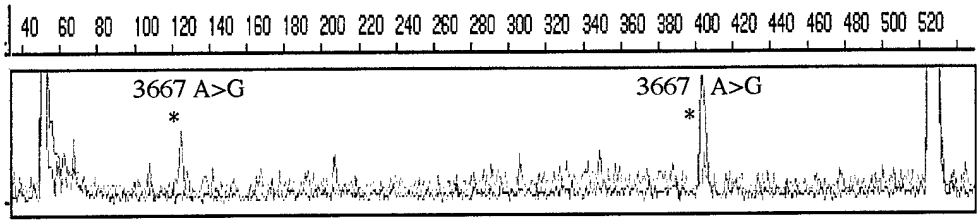
Exon 11
Sect. 2



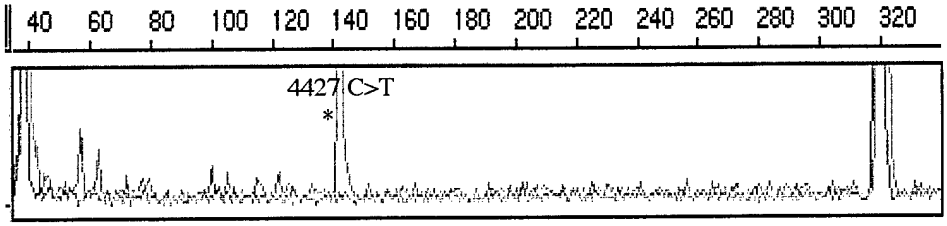
Exon 11
Sect. 4



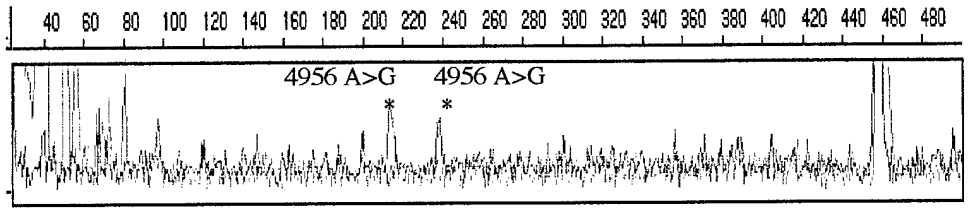
Exon 11
Sect. 5



Exon 11
Sect. 7

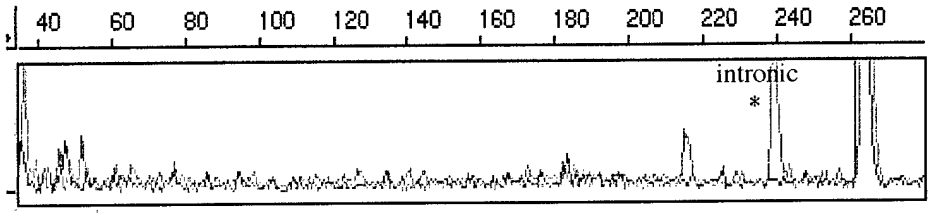


Exon 13

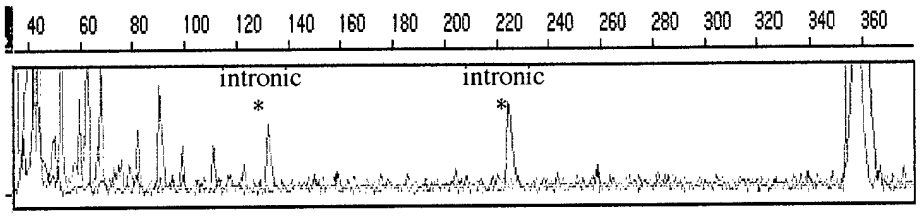


Exon 16

INDIVIDUAL #4

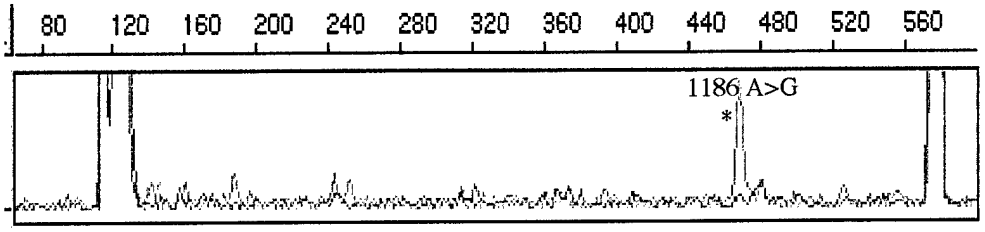


Exon 17

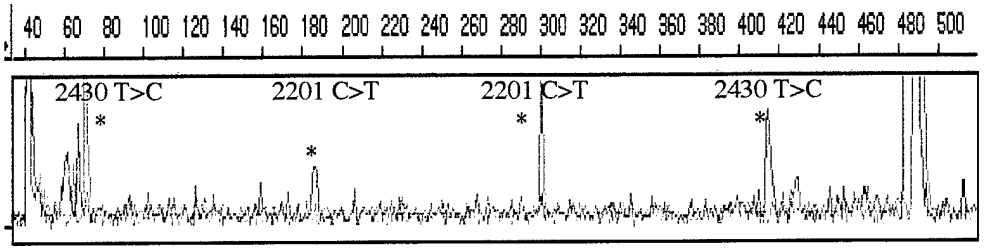


Exon 18

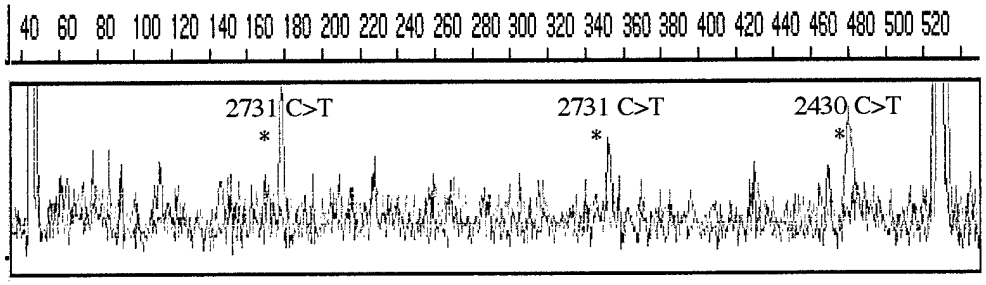
POLYMORPHISMS DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #5



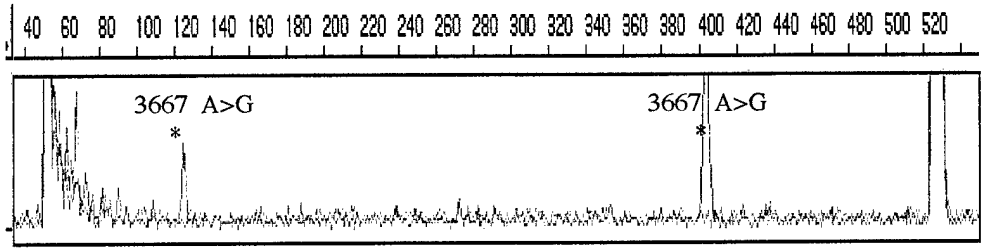
Exon 11
Sect. 2



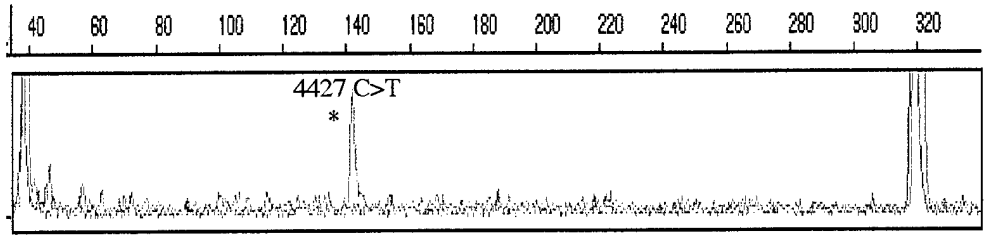
Exon 11
Sect. 4



Exon 11
Sect. 5

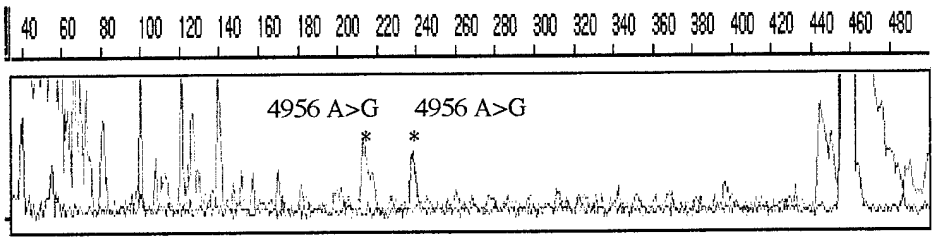


Exon 11
Sect. 7

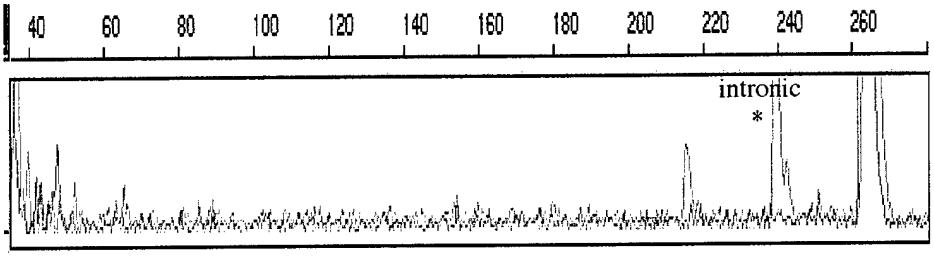


Exon 13

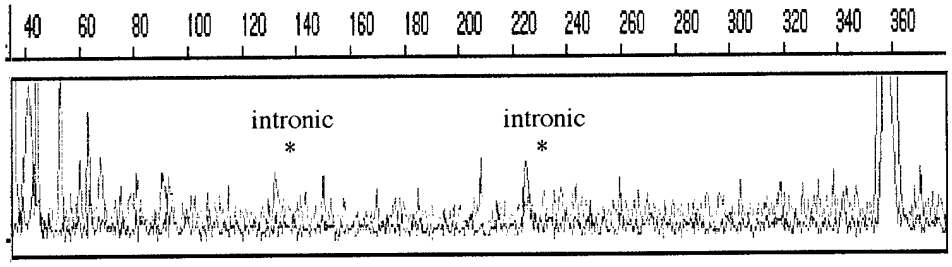
INDIVIDUAL #5



Exon 16



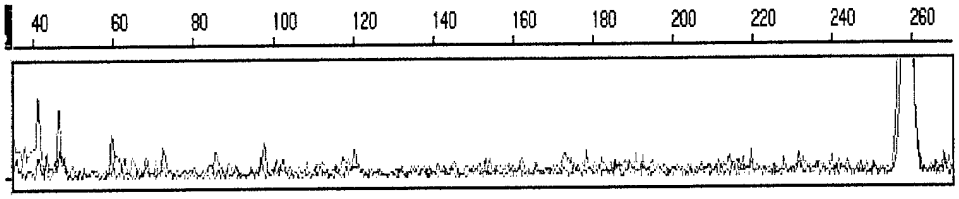
Exon 17



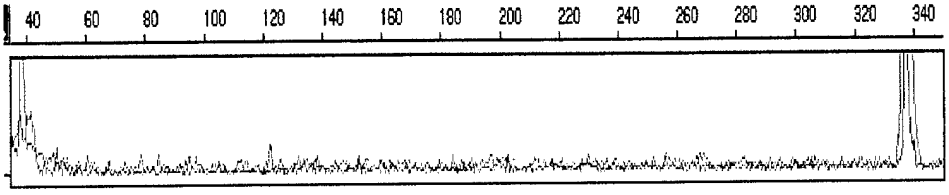
Exon 18

CEL I MUTATION DETECTION FOR *BRCA1* GENE FOR INDIVIDUAL #6

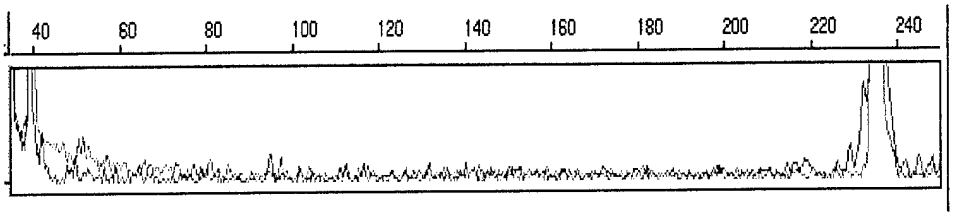
Electropherogram for each non-wizard prepped PCR product reacted with CEL I for Individual #6



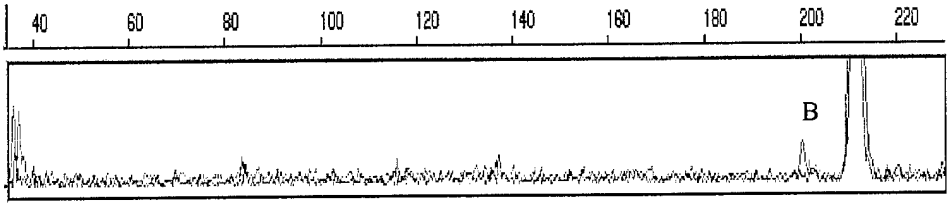
Exon 2



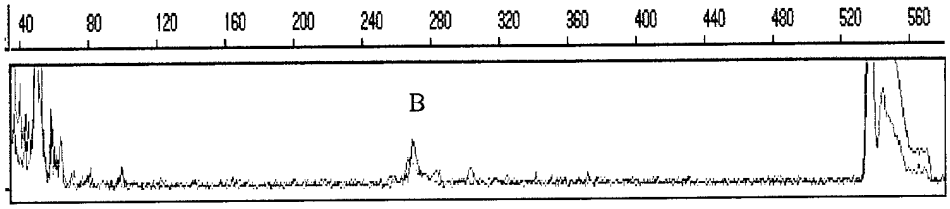
Exon 3



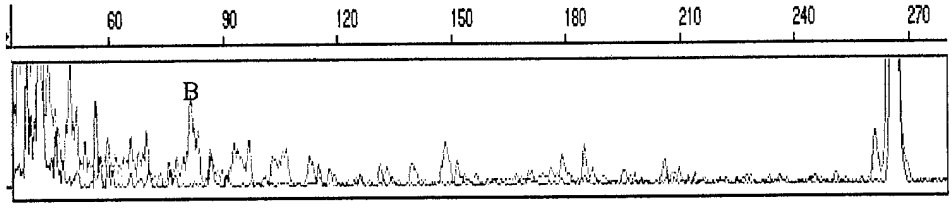
Exon 5



Exon 6

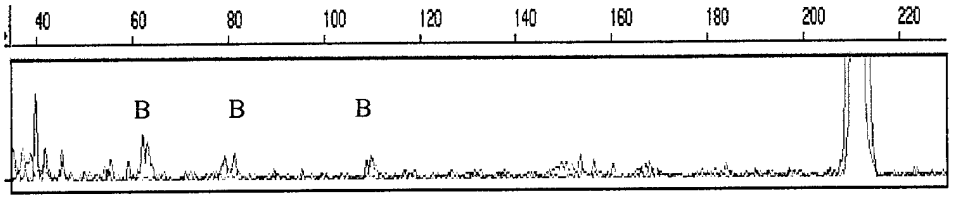


Exon 7

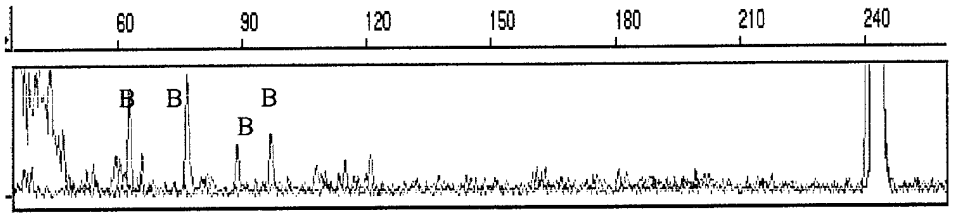


Exon 8

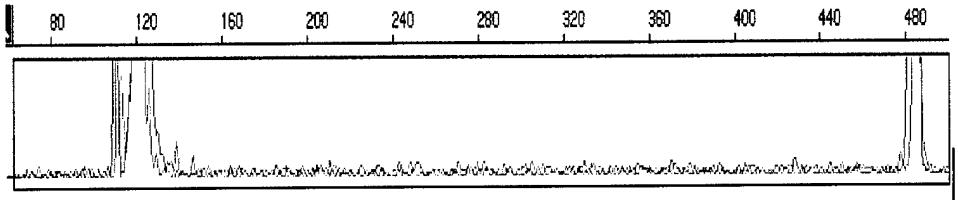
INDIVIDUAL #6



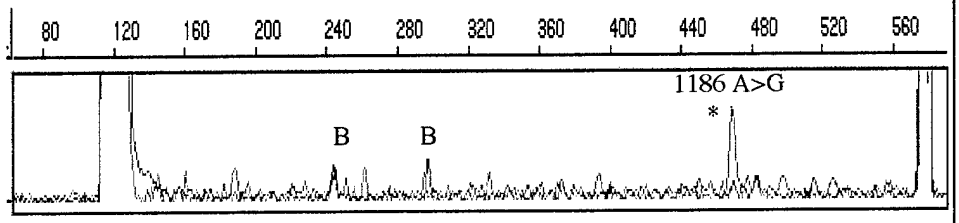
Exon 9



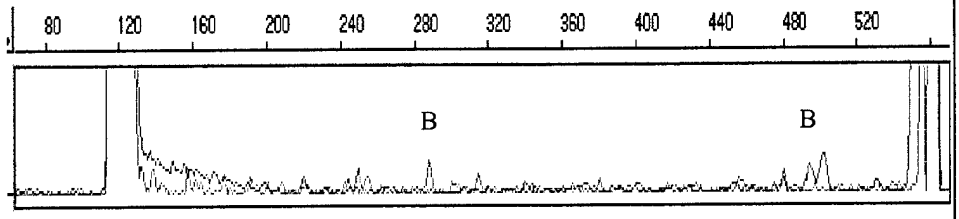
Exon 10



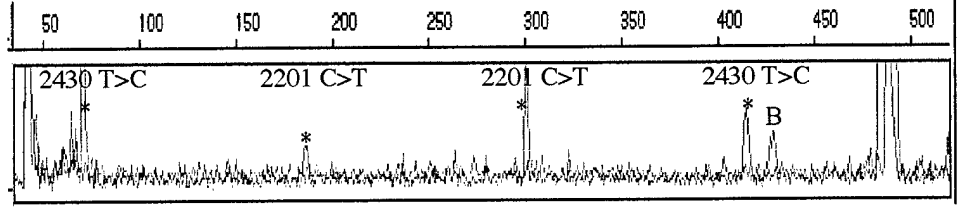
Exon 11
Sect. 1



Exon 11
Sect. 2

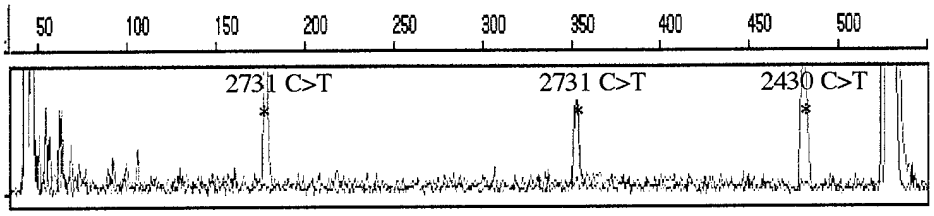


Exon 11
Sect. 3

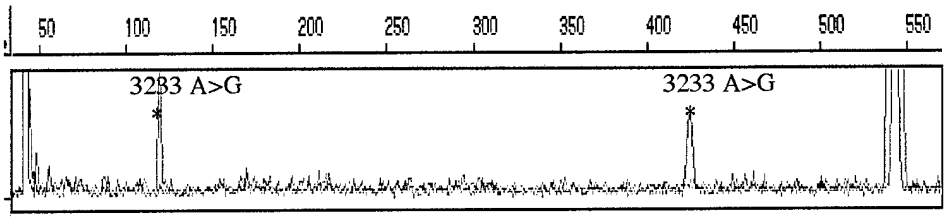


Exon 11
Sect. 4

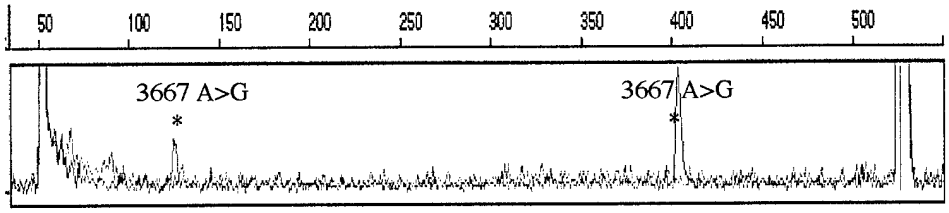
INDIVIDUAL #6



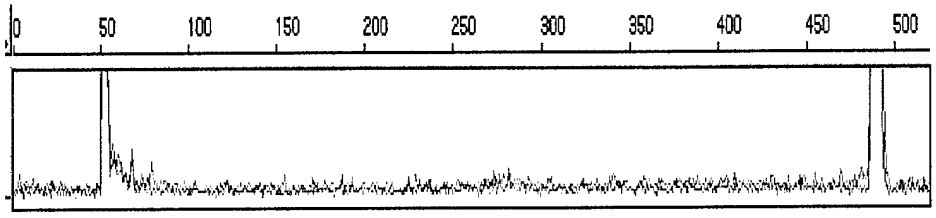
Exon 11
Sect. 5



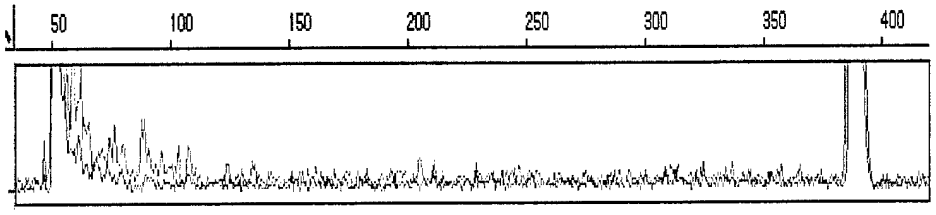
Exon 11
Sect. 6



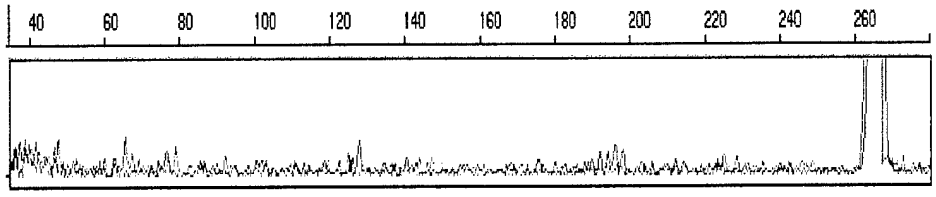
Exon 11
Sect. 7



Exon 11
Sect. 8

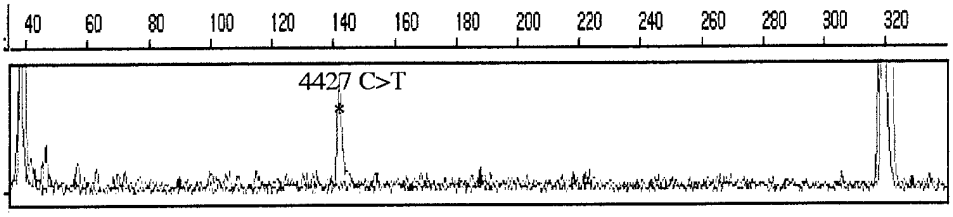


Exon 11
Sect. 9

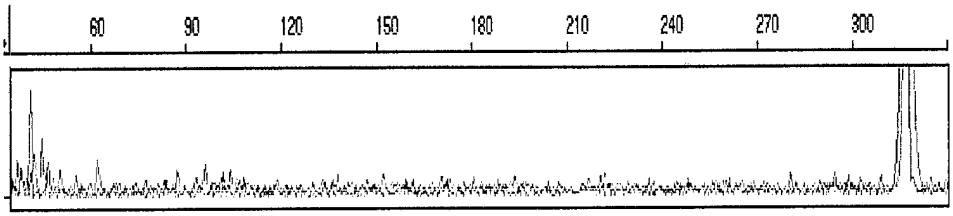


Exon 12

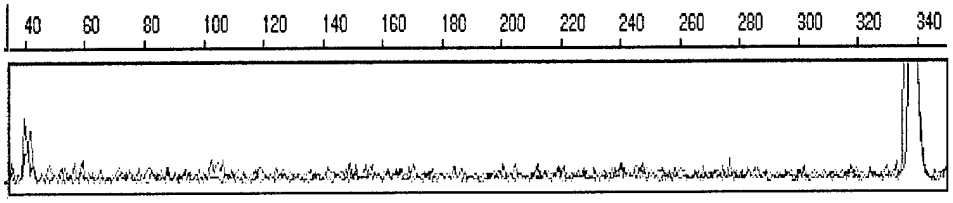
INDIVIDUAL #6



Exon 13



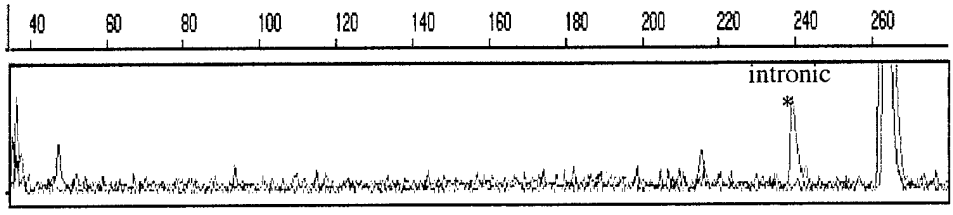
Exon 14



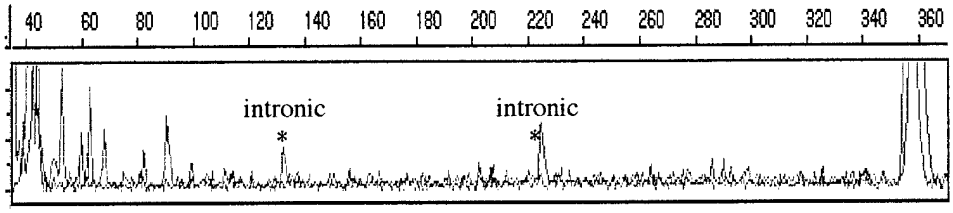
Exon 15

This particular gel lane contained no signal (no DNA) and has to be run again,
As of Sept 16, 1999 no conclusions have been made for ind. #6 exon 16

Exon 16

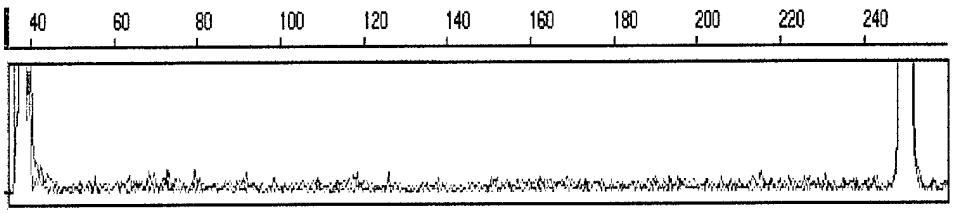


Exon 17

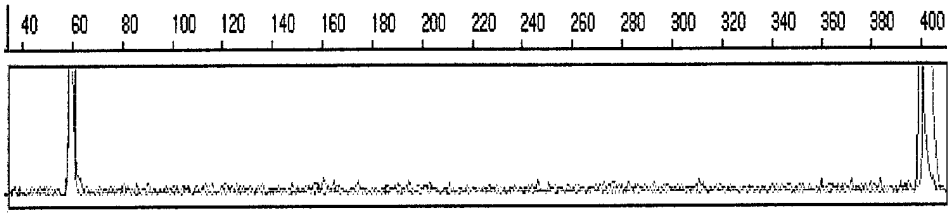


Exon 18

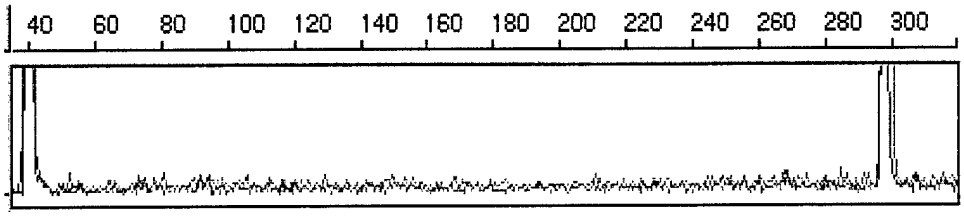
INDIVIDUAL #6



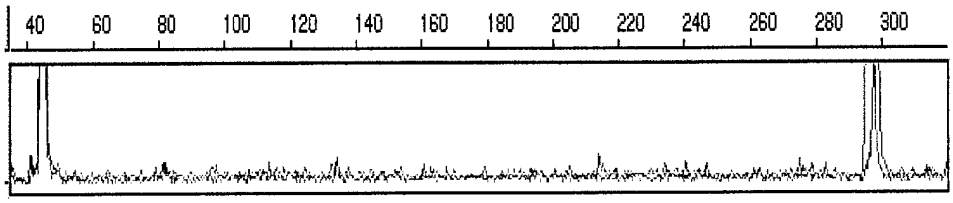
Exon 19



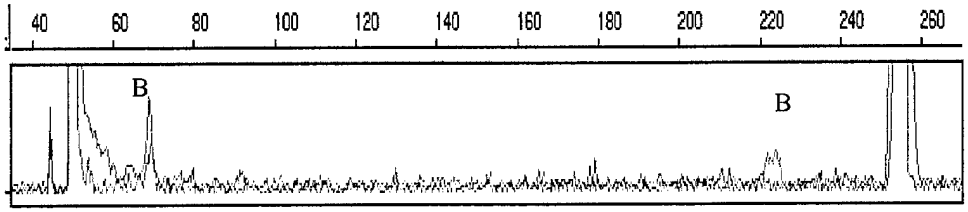
Exon 20



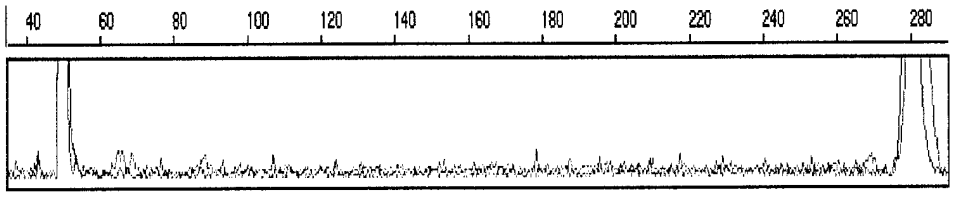
Exon 21



Exon 22

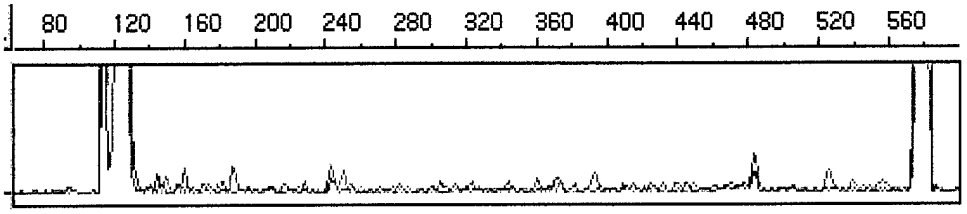


Exon 23

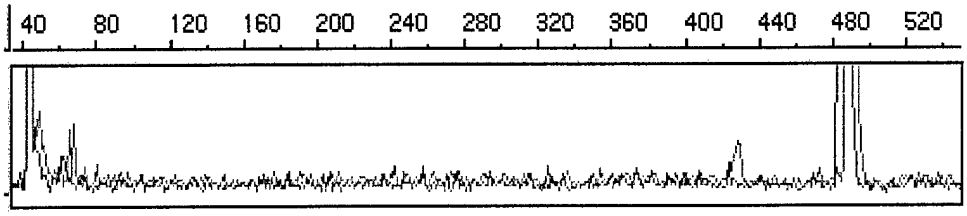


Exon 24

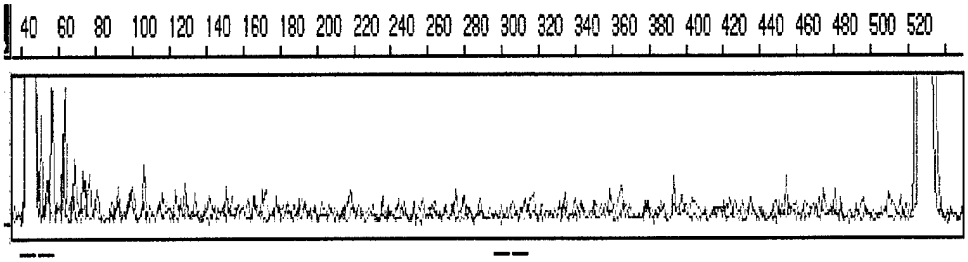
No polymorphisms detected for Individual #7
Exon 11(Sections 2,4,5,6,7) and Exons 13,16, 17,and 18



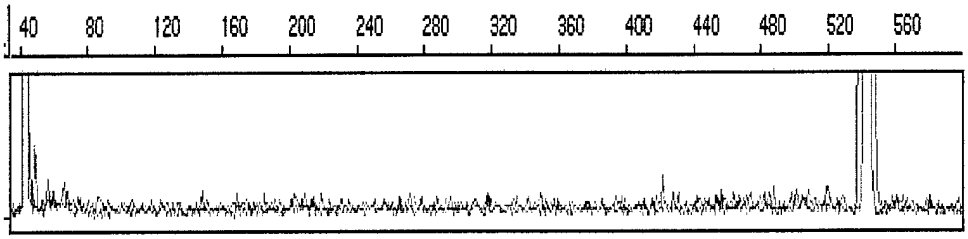
Exon 11
Sect. 2



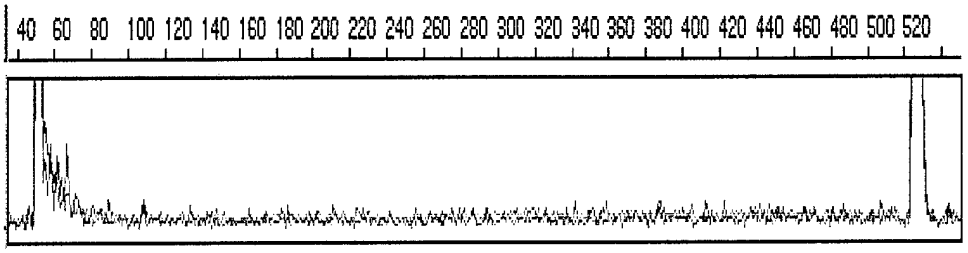
Exon 11
Sect. 4



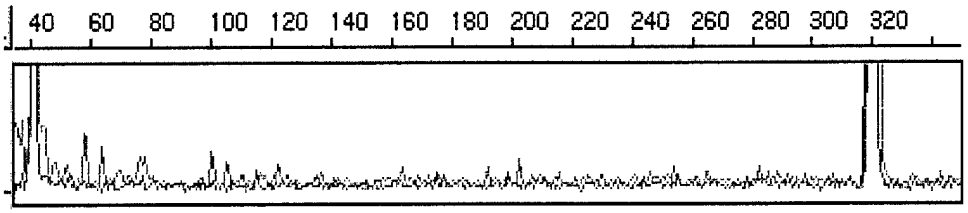
Exon 11
Sect. 5



Exon 11
Sect. 6

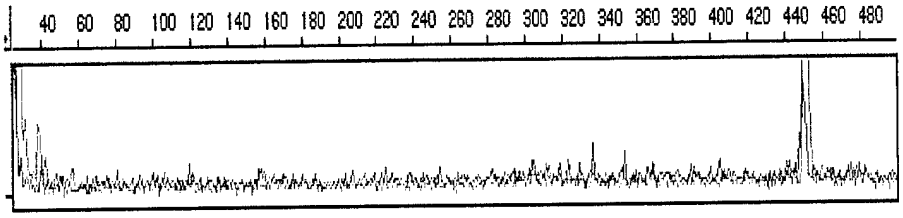


Exon 11
Sect. 7

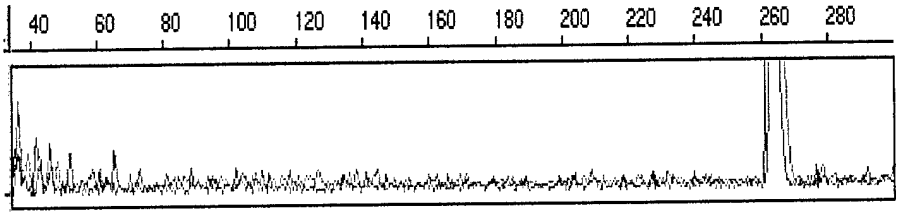


Exon 13

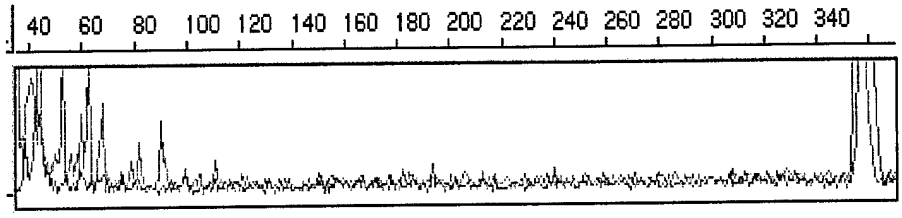
INDIVIDUAL #7



Exon 16

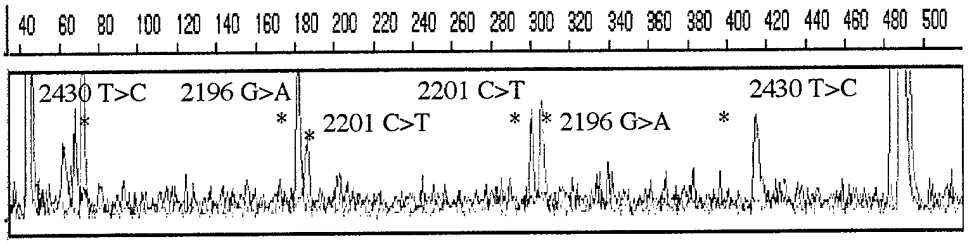


Exon 17

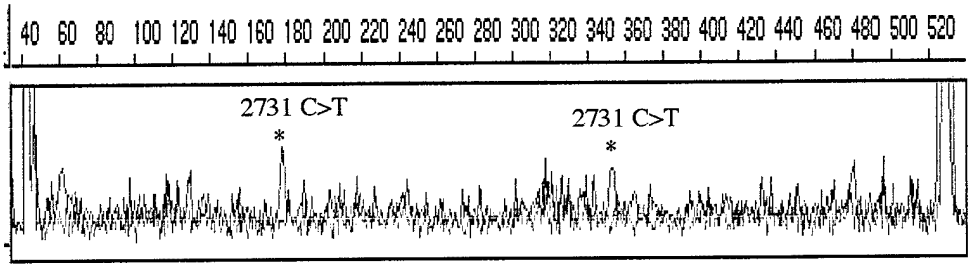


Exon 18

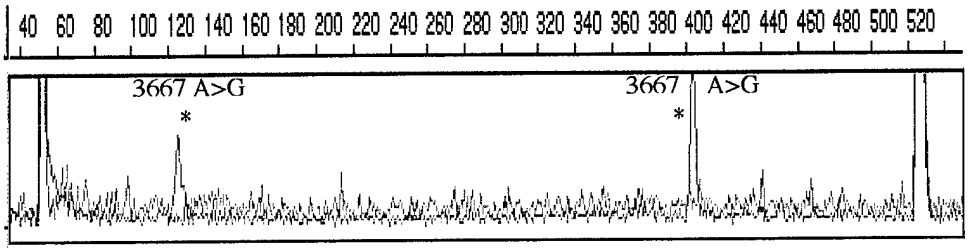
POLYMORPHISMS DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #8



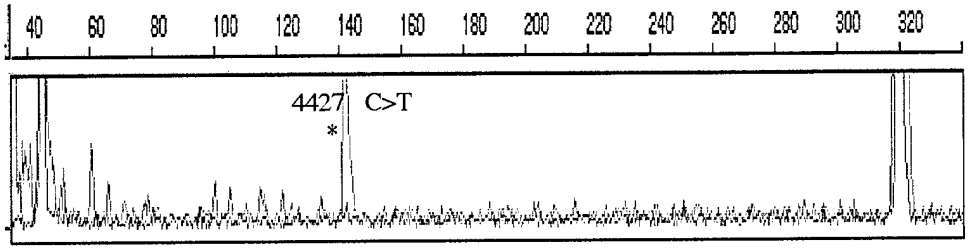
Exon 11
Sect. 4



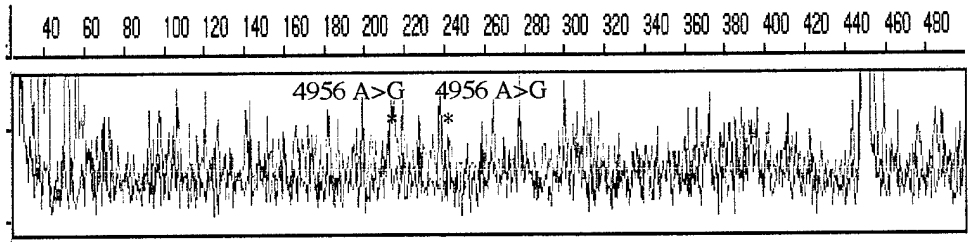
Exon 11
Sect. 5



Exon 11
Sect. 7



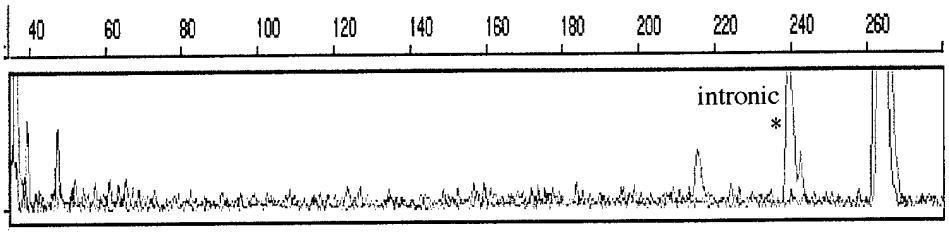
Exon 13



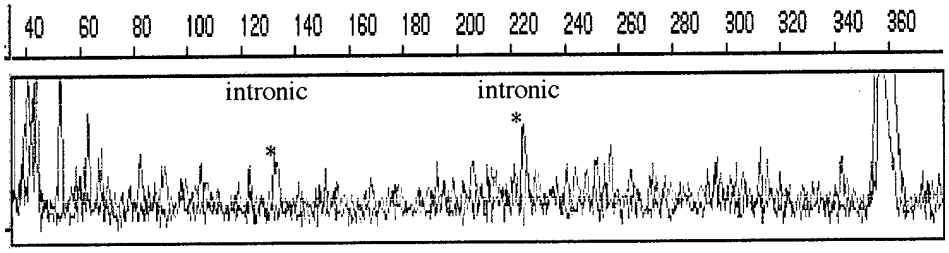
Exon 16

This particular electropherogram contained background noise and the amount of DNA seemed abnormally low, the evaluation of this particular sample is being re-run

INDIVIDUAL #8

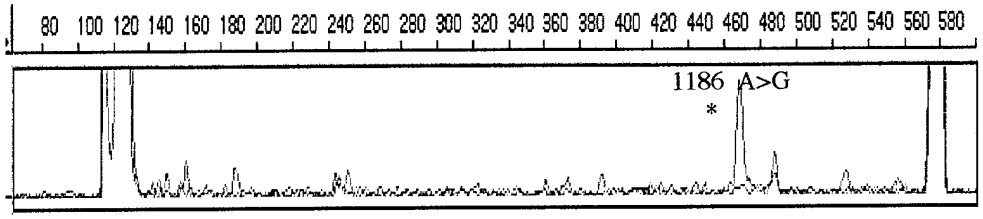


Exon 17

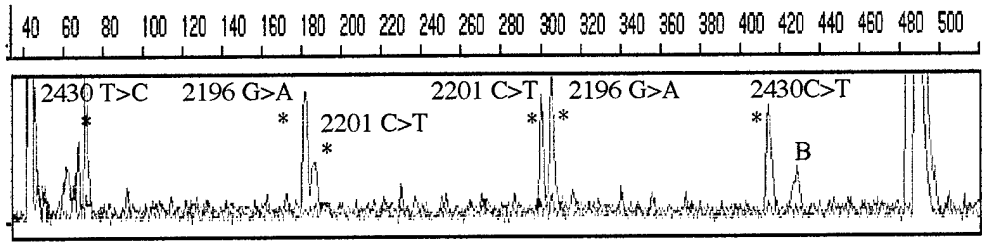


Exon 18

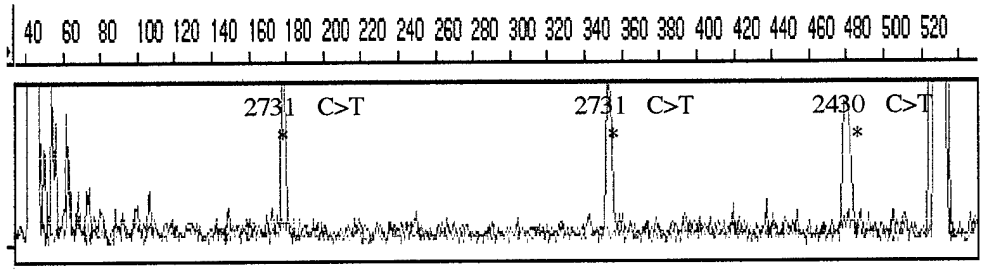
POLYMORPHISMS DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #9



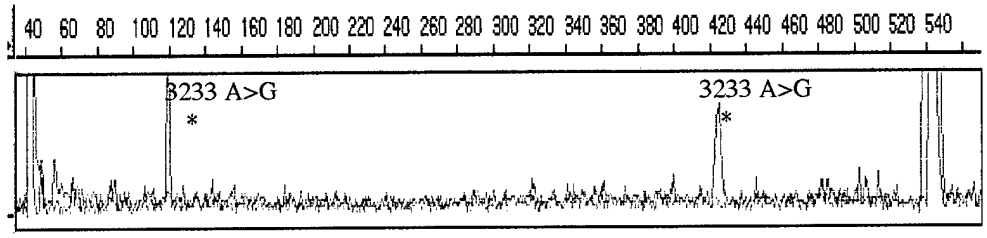
Exon 11
Sect. 2



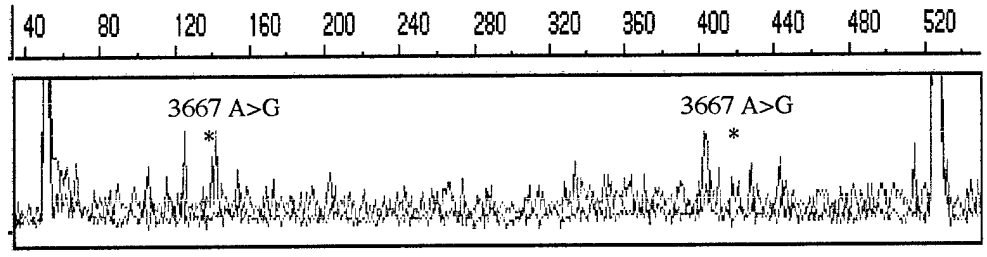
Exon 11
Sect. 4



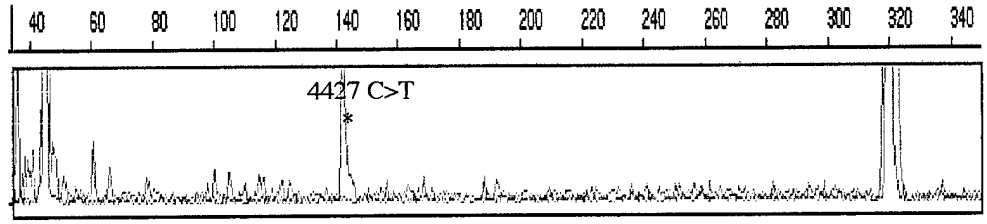
Exon 11
Sect. 5



Exon 11
Sect. 6

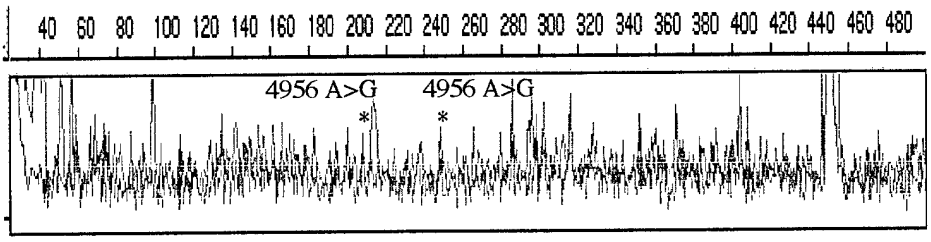


Exon 11
Sect. 7

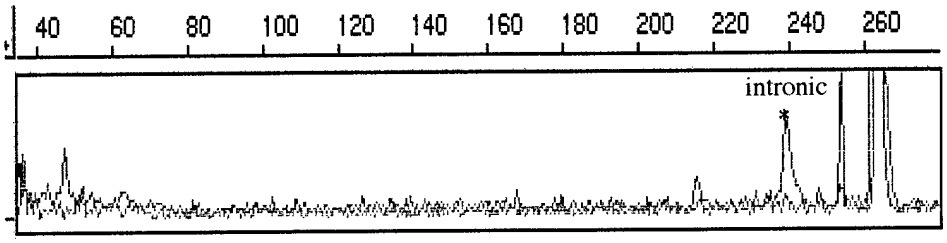


Exon 13

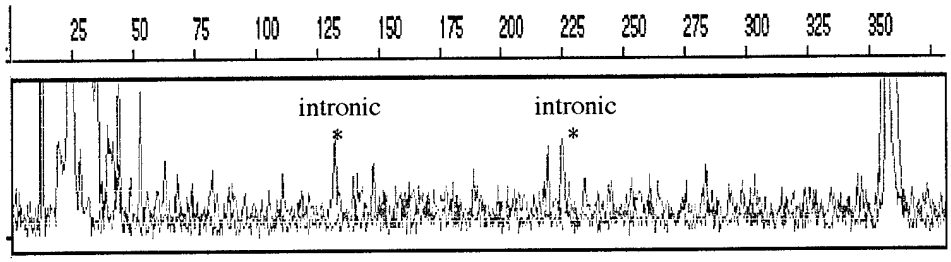
INDIVIDUAL #9



Exon 16

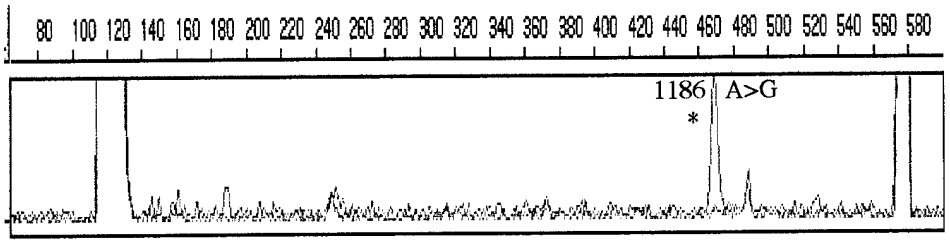


Exon 17



Exon 18

POLYMORPHISM DETECTED BY CEL I MUTATION DETECTION FOR INDIVIDUAL #10



Exon 11
Sect. 2

Specific Aim 2. Methods development: (i) To develop protocols to cover the whole *BRCA1* gene as six fragments of about 1 kilobase; (ii) To develop double-nested overlap-primer PCR approach to minimize the possibility of allelic loss in PCR; (iii) Multiplex and grid analysis of 300 basepairs long PCR products. (iv) Multi-kilobase PCR product mutation detection; (v) Using energy-transfer primers to enhance fluorescence.

Task 5: Months 1-3 Protocols to cover the whole *BRCA1* gene as six fragments of about 1 kilobase will be developed. RT-PCR will be tested. Bridging PCR will be tested.

This task was partially completed in that the limitations of current instrumentation and CEL I mutation detection technology was evaluated. It indicated that we should limit the target fragment size to about 600 bp for the ease of data analysis and gel electrophoresis. The 1000 bp fragments sometimes encounter more than one polymorphisms/mutations in the same region. Such situation causes an individual mutation detection signal to be diminished by the CEL I cuts at the second polymorphic site. The 600 bp target is a square root less likely to have two polymorphisms in the same DNA region. Therefore the task of parsing the *BRCA1* gene into six fragments of 1000 bp was deemed impractical at present.

Task 6: Months 4-5 Double-nested overlap-primer PCR approach to minimize the possibility of allelic loss in PCR will be tested.

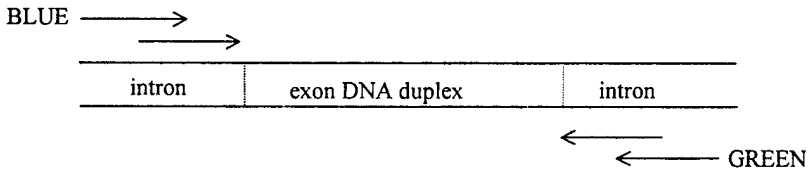
This approach was developed in a blind study of steroid sulfatase gene *ARSC* as well as the *BRCA1* gene. The results are described in details in the following manuscript:

Besack, D., Kulinski, J. A., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) Polymorphisms of the human steroid sulfatase gene. Manuscript in preparation, enclosed.

We chose the *ARSC* gene to perform this task because it is an important enzyme that may affect the onset of breast cancer as well as influence the efficacy of steroid chemotherapeutic drugs. Sulfation and desulfation are important reactions in the metabolism of many steroid hormones. Estrone, estradiol and dehydro-epiandrosterone (DHEA) circulate predominantly in the sulfated form and as such are not biologically active (i.e., do not bind target receptors). Furthermore, the sulfated forms of many steroid hormones exhibit half-lives up to ten-fold higher than the desulfated form. Biological "cycling" of sulfated/desulfated steroid hormones has been demonstrated. The sulfated moiety represents a readily accessible, yet biologically inactive, "storage" form for many steroid hormones whereby hydrolysis of the sulfate group (desulfation) regenerates the biologically active steroid. These observations suggest that sulfation and desulfation represent important reactions in the regulation of the biological activity of steroid hormones, and this regulatory system has become a target for chemotherapy of steroid hormone dependent tumors. ARSC, also known as steroid sulfatase (STS), catalyzes the desulfation of estrone-, 17β-estradiol-, and DHEA sulfate. As a first step to investigate whether functionally significant genetic

polymorphisms occur within ARSC, we used CEL I mutation detection to analyzed the ARSC structural gene of 100 normal persons for the possible presence of polymorphisms. Among 100 normal persons, we found one missense mutation at amino acid 6 from the N-terminal, Leu→Ile. Potentially, this mutation can lead to the individual expressing ARSC from an alternate translational start site. A second mutation is in two persons in the intron between exons 2 and 3. Lastly, a very common polymorphism was observed in the 3' UTR, observed in 38 persons. No mutation or polymorphism was observed in the promoter region of the 100 individuals. The lack of neutral polymorphisms in the ARSC gene of 100 individuals is surprising. Explanations include possible codon bias in its coding, and mRNA stability. Our results show that if there were frequent variations in ARSC enzyme levels among individuals, it is not at the level of protein sequence or promoter sequence.

We used this opportunity to test a new concept to screen mutations in patient samples in pairs. Secondly, we screened the gene for polymorphisms/mutations a second time with a new economical protocol in which only one pair of universal fluorescent PCR primers are used for the PCR of all the exons. The protocol is an adaptation of the nested PCR approach proposed for this task. Namely, The first round of PCR amplification is performed with unlabeled primers containing a common 5' 12-nt overhang. All forward primers contained the sequence 5' TGTGCGGTCCTC 3' and all reverse primers contained the sequence 5' TTGATCCTACAA 3'. A second round of PCR was then carried out using the same pair of universal fluorescent primers for all products. These were forward primer 5' *6-FAM* GCCAGAGTTGTGCGGTCCTC 3' and reverse primer 5' *TET* GCCCGACTTTGATCCTACAA 3'. The 3' 12 nucleotide of these primers contain the same 12 nucleotide sequence as the respective 12-nt overhang of the unlabeled primer. This approach is shown schematically below:



The results in this manuscript illustrate that the nested universal fluorescent primer method produces the same mutation detection efficacy as previous more expansive approach in which one new pair of fluorescent primers are used for each exon. The universal fluorescent primer approach also allow potential users of this technology to adapt their existing PCR primers to CEL I mutation detection protocol with minimal expense or delay needed for fluorescent primer synthesis.

Task 7: Months 6-8 Multiplex and grid analysis of 300 basepairs long PCR products will be tested.

This task was completed by examining 100 patients of *BRCA1*, with their DNA multiplexed at 5 samples per reaction. The first test was described in details in test (ii) of the manuscript :

Kulinski, J. A., Besack, D. , Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation, enclosed.

In this work, we showed that the CEL I mutation detection protocol allows rather simple conditions for multiplexing in mutation screening and can be easily adapted to increase the throughput of the user laboratories.

In the second test, we performed a multiplex assay of an exon of 25 persons. The PCR of the DNA of each person was amplified individually and evaluated by agarose gel electrophoresis. With the PCR products of the 25 people arranged in a 5x5 grid, we pooled the DNA of 5 people in columns as 5 pools, and then in rows as another 5 pools [Table I]. When CEL I cut a mismatch in lane C, representing a vertical pool of 5 samples, it appeared in another lane, H, representing another horizontal pool of 5 samples. The xy coordinate of the positive samples indicates that sample 13 contained the mismatch. The Gel image of the analysis is shown in the next figure. This figure shows a gel image of the analysis of the DNA products of a CEL I digestion. Lanes A-J represent the 10 DNA pools of Table I. The 156 nucleotide (nt) product of the CEL I incision at the mismatch is seen in lanes C and H, corresponding to a signal from sample 13 in Table I. The lanes denoted + and – are the positive and negative control lanes in which the sample DNA contains the same mismatch as in sample 13, or no mismatch, respectively.

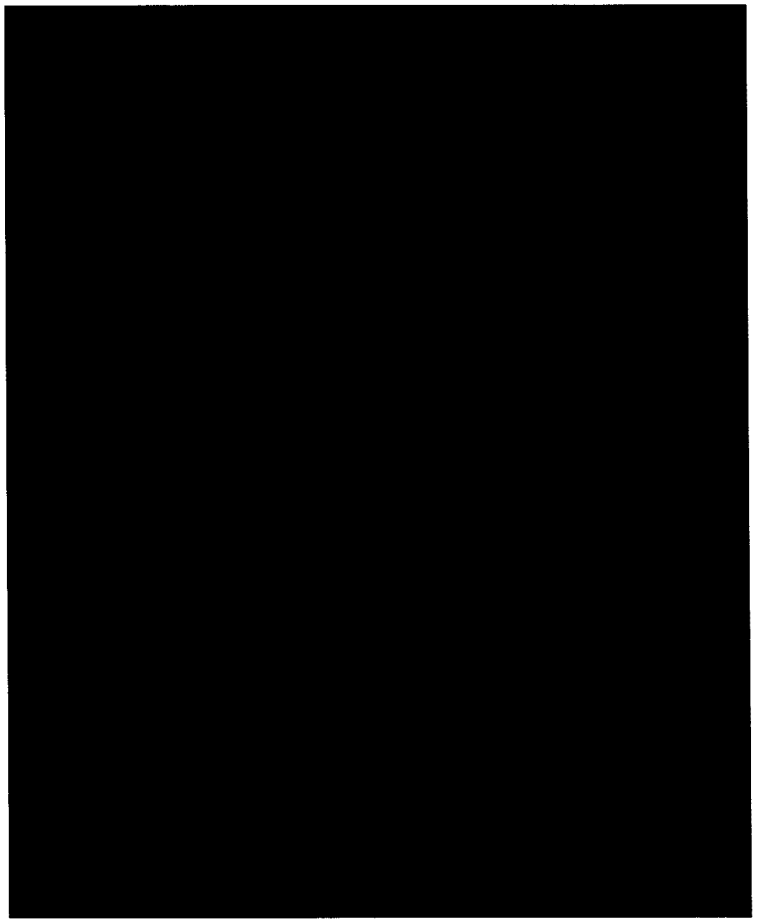
Table I: 5x5 multiplex grid

	F	G	H	I	J
A	1	2	3	4	5
B	6	7	8	9	10
C	11	12	13	14	15
D	16	17	18	19	20
E	21	22	23	24	25

25 samples tested in 10 lanes: 1 in 5 wizard prepped PCR products

BRCA1 Exon 5 T --> G mutation

A B C D E F G H I J + -



235 nt

156 nt

80 nt

Task 8: Months 9-10 Multi-kilobase PCR product mutation detection will be developed.

This task has been assigned to a user lab that we are supporting with CEL I. The collaborator is Dr. Evgeni V. Sokurenko, M.D., Ph.D. of the Dept. of Microbiology of the Univ. of Washington, Seattle, WA. He has modified the CEL I assay for the purpose of rapidly detecting a new mutation in the genome of a microbial pathogen and so far succeeded in detecting mutations in DNA fragments up to 70 Kbp in length. From examining his data, we believe that his approach can be pushed to become a whole genome scan of the microorganism. The CEL I system has allowed Dr. Sokurenko to obtain grant funds to continue this collaboration.

Task 9: Months 11-16 Energy-transfer primers to enhance fluorescence will be synthesized and tested.

This approach has assumed lower priority because of two reasons: First, the technology of energy transfer primers has been kept proprietary by Amersham Pharmacia Biotech and not made readily accessible. We have not had the time to develop alternate economical means of making energy transfer primers. Secondly, Neither Molecular Dynamics Inc. nor Hitachi Inc. has improved their fluorescence scanners from the two color ability to true four color ability with good color separation needed to practice CEL I mutation detection on their machines. Energy transfer primers were envisioned to benefit from the use of flat bed fluorescence scanner and must await the development of such scanners.

Tasks completed but not described in the original proposal: Months 11-24.

1. CEL I was purified, the amino acid sequence determined for 30% of the protein, the gene was cloned and sequenced, and the cDNA was used to expressed recombinant CEL I in a heterologous system. The expressed protein was partially purified and characterized. Parts of this work was described in the following two manuscripts:

   Yang, B., Wen, X., Oleykowski, C. A., Miller, C. G., Kulinski, J. A., Besack, D. A., and Yeung, A. T. (1999) Purification and Characterization of the CEL I Endonuclease that has high specificity for mismatch. Submitted to *J. Biol. Chem.* for publication.

   Kodali, N.S., Oleykowski, C.A., Kowalski, D., Yang, B., Miller, C.G., Besack, D.A., Kulinski, J.A., and Yeung, A.T. (1999) A Comparison of the CEL I Endonuclease with the Mung Bean Nuclease, two nucleases of the S1 superfamily. Submitted to *J. Biol. Chem.* for publication.

   The cloning and expression work, although completed, cannot be described in this report because of patenting requirements.

2. Samples of CEL I were sent to about 30 international laboratories to promote the establishment of the CEL I mutation detection method in these labs. This list of users is provided in the APPENDICES section. Many of these labs are interested in

screening genes related to breast cancer. Support of these laboratories have taken major effort on the part of Ms. Oleykowski in my lab. However, we consider getting the enzyme into the hands of users a major priority because some of the users will have the expertise to produce more data on how well CEL I mutation detection works in their hands. Other users were interested in using CEL I as a platform to develop second generation mutation detection protocols that our laboratory is not equipped to develop on our own.

KEY RESEARCH ACCOMPLISHMENTS:

1. We invented the CEL I enzymatic mutation detection assay. This robust assay uses a novel enzyme CEL I discovered and patented by this laboratory.
2. The CEL I mutation detection assay has been optimized.
3. We tested the CEL I mutation detection assay on the *BRCA1* gene of 10 patients.
4. We multiplex the CEL I mutation detection assay of a *BRCA1* exon of 100 patients.
5. The CEL I mutation detection assay was applied to the screening of the human steroid sulfatase gene ARSC of 100 persons. This gene is important to breast cancer.
6. To lower the cost of the assay, we developed a universal PCR fluorescent primers approach for the CEL I mutation detection assay.
7. To strengthen the CEL I nuclease system, we accomplished the purification, identification and peptide sequencing of the CEL I nuclease and the cloning, and expression of the CEL I nuclease gene.

REPORTABLE OUTCOMES:

Bibiography of publications:

1. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research, 26*, 4597-4602.
2. Oleykowski, C. A., Bronson Mullins, C. R., Chang, D. W., and Yeung, A. T. (1999) Incision at nucleotide insertions/deletions and basepair mismatches by the SP Nuclease of Spinach. *Biochemistry* **38**, 2200-2205.
3. Yang, B., Wen, X., Oleykowski, C. A., Miller, C. G., Kulinski, J. A., Besack, D. , and Yeung, A. T. (1999) Purification and Characterization of the CEL I Endonuclease that has high specificity for mismatch. Submitted to *J. Biol. Chem.* for publication.
4. Kodali, N.S., Oleykowski, C.A., Kowalski, D., Yang, B., Miller, C.G., Besack, D., Kulinski, J.A., and Yeung, A.T. (1999) A Comparison of the CEL I Endonuclease with the Mung Bean Nuclease, two nucleases of the S1 superfamily. Submitted to *J. Biol. Chem.* for publication.
5. Kulinski, J. A., Besack, D. , Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation.
6. Besack, D. , Kulinski, J. A., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) Polymorphisms of the human steroid sulfatase gene. Manuscript in preparation.

Patent:
U.S. patent number 5869245 (1999) "Mismatch endonuclease and its use in identifying mutations in targeted polynucleotide strands".

CONCLUSIONS:

The CEL I mutation detection assay has been optimized. The enzyme and the assay are shown to be robust. It is able to detect all mutations in a gene with a minimum of effort. Through our manuscripts, we have provided examples of how CEL I mutation detection assays can be multiplexed to increase throughput. The assay is at a point where commercialized distribution of the technology is feasible. This is made possible by the issuing of the patent on this technology, and by the purification and amino acid sequencing of CEL I, followed by the cloning and expression of recombinant CEL I. An extensive network of collaborations has been established to share this technology with scientists interested in mutations and cancer. Through these collaborations, the utility of the research supported by this grant may be further extended.

REFERENCES:

1. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, **26**, 4597-4602.
2. Castilla, L. H. Couch, F. J., Erdos, M. R., Hoskins, K. F., Calzone, K., Garber, J. E., Boyd, J., Lubin, M. B., Deshano, M. l., Brody, L. C., Collins, F., and Weber, B. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nature Genet.* **8**, 387-391, 1994.
3. Friedman, L. S., Ostermeyer, E. A., Csilla, I. S., Dowd, P., Lynch, E. D., Rowell, S. E., and King, M.-C. Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genet.* **8**, 399-404, 1994.
4. Ellison, J., Squires, G., Crutchfield, C., and Goldman, D. Detection of mutation and polymorphisms using fluorescence-based dideoxy fingerprinting (F-ddF). *Biotechniques* **17**, 742-753, 1994.
5. Brow, M. A., et. al., Mutation detection by Cleavase Fragment Length Polymorphism Analysis. *Focus* **18.1**, 2, 1996.
6. Hogervorst, F. B. L., Cornelis, R. S., Bout, M., van Vliet, M., Oosterwijk, J. C., Olmer, R., Bakker, B., Klijn, J. G. M., Vasen, H. F. A., Meijers-Heijboer, H., Menko, F. H., Cornelisse, C. J., den Dunnen, J. T., Devilee, P., and van Ommen, G.-J. B. Rapid detection of *BRCA1* mutation by the protein truncation test. *Nature Genetics* **10**, 208-213, 1995.
7. Lishanski, A., Ostrander, E. A., and Rine, J. Mutation detection by mismatch binding protein, MutS, in amplified DNA: application to the cystic fibrosis gene. *Proc. Nat. Acad. Sci. USA* **91**, 2674-2678, 1994.
8. Cotton, R. G. H., Rodrigues, N. R., and Campbell, R. D. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Nat. Acad. Sci.* **85**, 4397-4401, 1988.
9. Myers, R. M., Larin, Z., and Maniatis, T. detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**, 1242-1246, 1985.
10. Mashal, R. D., Koontz, J., and Sklar, J. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genetics* **9**, 177-183, 1995.
11. Miki, Y., Swenson, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Nennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonnin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Golgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994) . A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71.

APPENDICES:

International laboratories that received CEL I from us to facilitate their mutation detection studies and methods development:

Alec Morley
Professor and Head
Dept. of Haematology and Genetic Pathology
Flinders University and Medical Centre
Bedford Park, SA 5042
Australia

Bifeng Gao
University of Colorado Health Sciences Center
4200 E. 9th Ave. Box C-321
Room 1W19
Denver, CO 80262

Robert E. Ferrell
Department of Human Genetics
University of Pittsburgh
Pittsburgh, PA 15261

Doctor Brigitte Hermelin
Laboratoire de biologie moléculaire
Hopital Saint-Antoine
75012 PARIS
FRANCE

Charles W. Richard MD, PhD
Asst Professor of Psychiatry and Human Genetics
WPIC Rm 1445
University of Pittsburgh Medical Center
3811 O'Hara Street
Pittsburgh, PA 15213-2593

Dr. Chongjun Xu
Hyseq Inc.
675 Almanor Ave.
Sunnyvale, CA 94086

Zhi-Hao Qiu Ph.D.
Research Scientist
Clontech Laboratories, Inc.
1020 E. Meadow Circle
Palo Alto, CA 94303

David F. Bishop, Ph.D.
Professor of Human Genetics
Dept. of Human Genetics, Box 1498
Mount Sinai School of Medicine
1425 Madison Avenue, NY,NY 10029

David Gopurenko
D.Gopurenko@mailbox.gu.edu.au
Ph. D. student. Australia (received instructions from us to purify CEL I.)

Dr D Hills
Roslin Institute
Midlothian
EH25 9PS
UK

David Hornby
Department of Molecular Biology and Biotechnology,
University of Sheffield
Western Bank Sheffield S10 2TN, UK.

David L. Gillespie:
Dept. of Biochemistry
University of New Mexico

Duncan Clark
DNAmp Ltd.

Evgeni V. Sokurenko, MD, Ph.D.
Research Assistant Professor
Dept. of Microbiology
Univ. of Washington
Box 357242
Seattle, WA 98195-7242

Frank-Ulrich Gast, PhD
Justus-Liebig-Universitaet Giessen
Institut fuer Biochemie
Heinrich-Buff-Ring 58
D-35392 Giessen

Graham Taylor PhD MRCPath
Scientific Director, Senior Associate Lecturer
Regional DNA Laboratory
St James's University Hospital
Leeds LS9 7TF
United Kingdom

Dr I R Graham
Research Officer
School of Biological Sciences
Division of Biochemistry
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK

Jean - Pierre de Villartay
INSERM U429, Pavillon Kirmisson
Hpital Necker-Enfants Malades
149 rue de Svres
75015 Paris, France

Jean-Laurent Casanova <casanova@necker.fr

Marcia Fournier
Dana-Farber Cancer Institute
44 Binney St
Boston, MA 02115

Mark Williams
DuPont AG Biotechnology
Stine-Haskell Research Center 210N/253
1090 Elkton Road
Newark, DE 19711-0030

Michael Bausher
USDA-ARS
2120 Camden Rd
Orlando, FL 32803

Niels Rüdiger, PhD
Department of Clinical Biochemistry
University Hospital of Aarhus
Skejby Hospital
DK-8200 Aarhus N
Denmark

Oliver Mueller
Max-Planck-Institut fuer molekulare Physiologie
Rheinlanddamm 201
44139 Dortmund
Germany
Paul M. Lizardi, PhD
Associate Professor
Dept. of Pathology
Yale School of Medicine
310 Cedar Street, New Haven CT 06510

Phillip R. Musich, Ph.D.
Department of Biochemistry and Molecular Biology
J.H. Quillen College of Medicine
East Tennessee State University
Johnson City, TN 37614-0581

Park, Sang-Ryoul, Ph.D.
730 Hilton
Clinical Biochemistry and Immunology
Mayo Clinic
Rochester, MN 55905

Thierry Frebourg, M.D, PhD
Laboratoire de Génétique Moléculaire
Faculté de Médecine et de Pharmacie
22 Boulevard Gambetta, 76183 Rouen Cedex, France

Thomas Jansen
IMD
Institut fuer molekularbiologische Diagnostik
Endenicher Allee 15
D-53115 Bonn
Germany

Warren Glaab
Merck Research Laboratories
WP45-320
West Point, PA 19486

Xinghua Pan,
Boyer Center for Molecular Medicine
Department of Genetics
Yale School of Medicine

For FINAL REPORT:

Bibiography of publications:

1. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research, 26*, 4597-4602.
2. Oleykowski, C. A., Bronson Mullins, C. R., Chang, D. W., and Yeung, A. T. (1999) Incision at nucleotide insertions/deletions and basepair mismatches by the SP Nuclease of Spinach. *Biochemistry* **38**, 2200-2205.
3. Yang, B., Wen, X., Oleykowski, C. A., Miller, C. G., Kulinski, J. A., Besack, D., and Yeung, A. T. (1999) Purification and Characterization of the CEL I Endonuclease that has high specificity for mismatch. Submitted to *J. Biol. Chem.* for publication.
4. Kodali, N.S., Oleykowski, C.A., Kowalski, D., Yang, B., Miller, C.G., Besack, D., Kulinski, J.A., and Yeung, A.T. (1999) A Comparison of the CEL I Endonuclease with the Mung Bean Nuclease, two nucleases of the S1 superfamily. Submitted to *J. Biol. Chem.* for publication.
5. Besack, D., Kulinski, J. A., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) Polymorphisms of the human steroid sulfatase gene. Manuscript in preparation, draft enclosed.
6. Kulinski, J. A., Besack, D., Oleykowski, C. A., Yang, B., Miller, C. G., and Yeung, A. T. (1999) The CEL I enzymatic mutation detection assay. Manuscript in preparation, draft enclosed.

Patent:

U.S. patent number 5869245 (1999) "Mismatch endonuclease and its use in identifying mutations in targeted polynucleotide strands".

Meeting Abstract:

Meeting of the Association of Biomolecular Resources Facilities (ABRF)
March 1998, San Diego, Calif.

CHARACTERIZATION OF THE SITE OF CEL I MISMATCH ENDONUCLEASE INCISION BY MALDI-TOF. Anthony T. Yeung and Catherine A. Oleykowski. Fox Chase Cancer Center, Philadelphia PA 19111

Celery contains a nuclease, CEL I, that is highly specific for insertional/deletional DNA loop lesions and mismatches. The exact bond of DNA incision of the CEL I nuclease is the subject of this investigation. While DNA sequencing gel analysis showed that the phosphodiester bond broken in the DNA incision is on the 3' side of the mismatch nucleotide, the proof whether the incision produces a 3'-OH group is absent. In this study, we used MALDI-TOF mass-spectrometry to measure the exact mass of the product of the CEL I incision. From this study, it can be concluded that the CEL I incision produces a 3'-OH group and a 5' PO_4 group at the mismatch site. The presence of a 3'-OH group will allow DNA repair to continue by the binding of a DNA

polymerase. Had a 3'-PO_4 group been present, a 3' phosphatase would have been required in the repair mechanism. CEL I also exhibit an exonuclease activity that is specific for the DNA duplex 3' and 5' termini on the 3' side of the mismatch incision site. The mismatch endonuclease activity and the exonuclease activity appears to be tightly coupled.

List of personnel:

Receiving salary from this grant:

Anthony T. Yeung, Ph.D. Principal Investigator, 10% FTE.

Catherine A. Oleykowski, M.S. Research Assistant, 100% FTE. She was instrumental to the development of the CEL I nuclease mutation detection method. She purified the CEL I enzyme, and partially optimized the mutation detection assay. She assisted Dr. Yeung in the preparation and defense of the CEL I mismatch endonuclease patent. She is in charge of the time consuming interaction and support of the two dozen CEL I international user laboratories. She trained David Besack, Joanne Kulinski, and several users of our institution in using the CEL I system.

Not receiving salary from this grant:

David Besack, B.S. Research Technician (He received training from Ms. Oleykowski and did the optimization of the CEL I Genescan protocol and the Steroid sulfatase work).

Joanne Kulinski, B.S. Research Technician (She was trained by Ms. Oleykowski and did the blind study evaluation of the use of CEL I on 100 patients, and the two multiplex studies. Thereafter, she was no longer directly involved in this project, but proceeded to clone and express the CEL I gene successfully).

Bing Yang, Ph.D. Post-doctoral fellow (his participation in this project is to providing some of the purified CEL I nuclease, and instruction to Joanne Kulinski on the cloning and expression systems).

Nucleic Acids Research, 1998, Vol. 26, No. 20 4597–4602

Mutation detection using a novel plant endonuclease

Catherine A. Oleykowski, Colleen R. Bronson Mullins, Andrew K. Godwin and Anthony T. Yeung*

Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

Received July 8, 1998; Revised and Accepted August 25, 1998

ABSTRACT

We have discovered a useful new reagent for mutation detection, a novel nuclease CEL I from celery. It is specific for DNA distortions and mismatches from pH 6 to 9. Incision is on the 3′-side of the mismatch site in one of the two DNA strands in a heteroduplex. CEL I-like nucleases are found in many plants. We report here that a simple method of enzyme mutation detection using CEL I can efficiently identify mutations and polymorphisms. To illustrate the efficacy of this approach, the exons of the *BRCA1* gene were amplified by PCR using primers 5′-labeled with fluorescent dyes of two colors. The PCR products were annealed to form heteroduplexes and subjected to CEL I incision. In GeneScan analyses with a PE Applied Biosystems automated DNA sequencer, two independent incision events, one in each strand, produce truncated fragments of two colors that complement each other to confirm the position of the mismatch. CEL I can detect 100% of the sequence variants present, including deletions, insertions and missense alterations. Our results indicate that CEL I mutation detection is a highly sensitive method for detecting both polymorphisms and disease-causing mutations in DNA fragments as long as 1120 bp in length.

INTRODUCTION

Single-stranded nucleases such as S1 and mung bean nuclease nick DNA at single-stranded regions (1–3). However, the acid pH optima of these nucleases lead to DNA unwinding at A+T-rich regions and result in non-specific DNA degradation. For example, S1 nuclease was found not to cleave DNA at single base mismatches (4). The efficiency of mung bean nuclease at nicking supercoiled DNA is five orders of magnitude higher at pH 5 than at pH 8 (5). At neutral pH, a high concentration of mung bean nuclease is necessary to act on double-stranded DNA, mainly at A+T-rich regions (3). In this report, we show that celery and many plants possess novel endonucleases, characterized by neutral pH optima, that detect destabilized regions of DNA helices, such as at the site of a mismatch. The celery enzyme was named CEL I. The mismatch specificity of CEL I at neutral pH has enabled development of a highly effective and user-friendly mutation detection assay. We illustrate this CEL I method by detection of mutations and polymorphisms of the *BRCA1* gene of a number of women affected with either breast and/or ovarian cancer and reporting a family history of these diseases.

MATERIALS AND METHODS

Preparation of plant extracts

Various plant tissues were homogenized in a Waring blender at 4°C and adjusted with a 10× solution to give the composition of buffer A [0.1 M Tris–HCl, pH 7.7, 10 μM phenylmethanesulfonyl fluoride (PMSF)]. The extracts were stored at –70°C. Equivalent data were obtained when the tissues were frozen in liquid nitrogen, ground to a powder with a mortar and pestle and then extracted with buffer A on ice.

Purification of CEL I

Celery stalks (7 kg) were extracted at 4°C with a juicer and adjusted with a 10× solution to give the composition of buffer A. The extract was concentrated with a 20–70% saturated ammonium sulfate precipitation step. The final pellet was dissolved in 250 ml buffer A and dialyzed against 0.5 M KCl in buffer A. The solution was incubated with 10 ml concanavalin A–Sepharose resin (Sigma) overnight at 4°C. The slurry was packed into a 2.5 cm diameter column and washed with 0.5 M KCl in buffer A. Bound CEL I was eluted with 90 ml 0.3 M α-D+-mannose, 0.5 M KCl in buffer A at 65°C. CEL I was dialyzed against buffer B (25 mM potassium phosphate, 10 μM PMSF, pH 7.0) and applied to a 100 ml phosphocellulose P-11 column that had been equilibrated in buffer B. The bound enzyme was eluted with a linear gradient of KCl in buffer B. The peak of CEL I activity was next concentrated by dialysis against saturated ammonium sulfate. The enzyme precipitate was dialyzed against buffer C (50 mM Tris–HCl, pH 7.8, 0.2 M KCl, 10 μM PMSF, 1 mM $ZnCl_2$) and fractionated by size exclusion chromatography on a Superose 12 FPLC column in the same buffer. The center of the CEL I activity peak from this step was used as the purified CEL I in this study. Protein concentrations of the samples were determined by the Bicinchoninic acid protein assay (Pierce).

Preparation of mismatch-containing heteroduplexes

The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer in the Fannie E.Rippel Biotechnology Facility of our Institution and purified using a denaturing PAGE gel in the

*To whom correspondence should be addressed. Tel: +1 215 728 2488; Fax: +1 215 728 3647; Email: at_yeung@fccc.edu

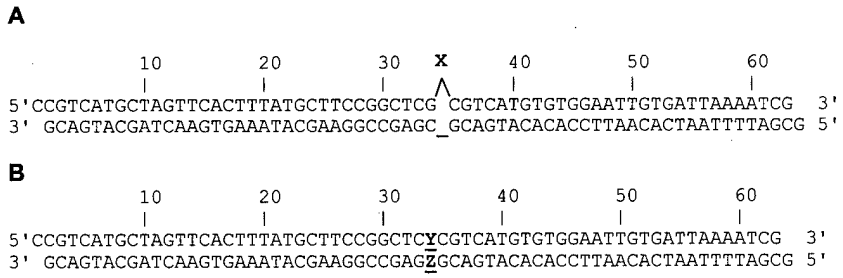


Figure 1. Design of the heteroduplexes containing base substitutions or DNA insertions. (**A**) Substrates with extrahelical DNA loop; (**B**) substrates with base substitution. Oligonucleotides containing variations of the nucleotides X, Y and Z were used to assemble all the permutations of mispaired substrates.

presence of 7 M urea at 50°C. DNA heteroduplex substrates of ~64 bp long containing mismatched base pairs or DNA loops (Fig. 1) were constructed by annealing partially complementary oligonucleotides. The single-stranded oligonucleotides were labeled at the 5′-termini with T4 polynucleotide kinase and [γ-^{32}P]ATP prior to annealing with an unlabeled oligonucleotide. After annealing, all the substrates were made blunt-ended by the fill-in reaction of DNA polymerase I Klenow fragment using dCTP and dGTP and purified by non-denaturing PAGE as described (6) without exposure to UV light. DNA was eluted from the gel slices using an AMICON model 57005 electroeluter.

Mismatch endonuclease assay

5′-^{32}P-labeled substrates (50–100 fmol) were incubated with CEL I preparation in buffer D (20 mM Tris–HCl, pH 7.4, 25 mM KCl, 10 mM $MgCl_2$) for 30 min at 37 or 45°C in 20 µl reactions. Taq DNA polymerase (0.5–2.5 U) (Perkin Elmer) was added to each reaction where indicated. The presence of dNTP is not necessary for DNA polymerase to stimulate CEL I turnover. Ten micromolar dNTP was included only in the reactions of Figure 3A to illustrate a form of nick translation that may result when dNTP is present. The reaction was terminated by adding 10 µl 1.5% SDS, 47 mM EDTA and 75% formamide plus tracking dyes, and analyzed on a denaturing 15% PAGE gel in 7 M urea run at 50°C. Autoradiography was used to visualize the radioactive bands. Chemical DNA sequencing ladders were included as size markers as previously described (6).

Sample ascertainment

As part of a Fox Chase Cancer Center (FCCC) Institutional Review Board approved protocol, peripheral blood samples were obtained from consenting affected high risk family members through the Margaret Dyson/Family Risk Assessment Program (FRAP). Individuals participating in FRAP have agreed to allow their samples to be used for a wide range of research purposes, including screening for mutations in candidate cancer predisposing genes, such as *BRCA1* (7). The participating individuals had previously been screened for *BRCA1* mutations by the Clinical Genetic Testing Laboratory at FCCC and were screened for sequence alterations by CEL I mutation detection in this study in a blind fashion.

DNA templates for *BRCA1* mutation analysis

Twenty five pairs of PCR primers specific for 22 coding exons in *BRCA1* were synthesized with 6-FAM dye (blue) at the 5′-end of each forward primer and with TET dye (green) at the 5′-end of each reverse primer. PCR was performed in a reaction volume of 20 µl containing 100 ng genomic DNA as template, 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 1 µM both forward and reverse primer, 60 µM each deoxyribonucleotide triphosphate, 5% dimethyl sulfoxide (DMSO) and 0.5 U Taq DNA polymerase. After an initial denaturation step at 94°C, the DNA was amplified through 20 cycles consisting of 5 s denaturation at 94°C, 1 min annealing at 65°C, decreasing by 0.5°C/cycle, and 1 min extension at 72°C. The samples were then subjected to an additional 30 cycles consisting of 5 s denaturation at 94°C, 1 min annealing at 55°C and 1 min extension at 72°C, with a final extension for 5 min at 72°C. The PCR reactions were purified using Wizard PCR Preps (Promega). The sizes of the DNA fragments generated by PCR ranged from 211 to 1120 bp.

CEL I mutation detection

Aliquots of 50–100 ng Wizard Prep processed DNA was heated to 94°C in buffer D and cooled to room temperature to form heteroduplexes. The heteroduplexes were incubated in 20 µl buffer D with 0.1 µl purified CEL I (0.01 µg) and 0.5 U Taq DNA polymerase at 45°C for 30 min. No dNTP was added. The reactions were stopped with 1 mM *o*-phenanthroline and incubated for an additional 10 min at 45°C. The samples were processed through a Centricep column (Princeton Separations) and dried in a SpeedVac. One microliter of ABI loading buffer (25 mM EDTA, pH 8.0, 50 mg/ml Blue Dextran), 4 µl deionized formamide and 0.5 µl TAMRA internal lane standard were added to the dried DNA pellet. The sample was heated at 90°C for 2 min, loaded onto a denaturing 34 cm well-to-read 4.25% polyacrylamide gel and analyzed on an ABI 373 Sequencer using GeneScan 672 Software (Perkin Elmer). Since the heteroduplexes were labeled with a different color on each strand, the mismatch-specific DNA nicking in each strand gave DNA fragments of two colors and different sizes that independently and complementarily pinpointed the mutation or polymorphism. All mutations and polymorphisms detected were confirmed by automated sequencing.

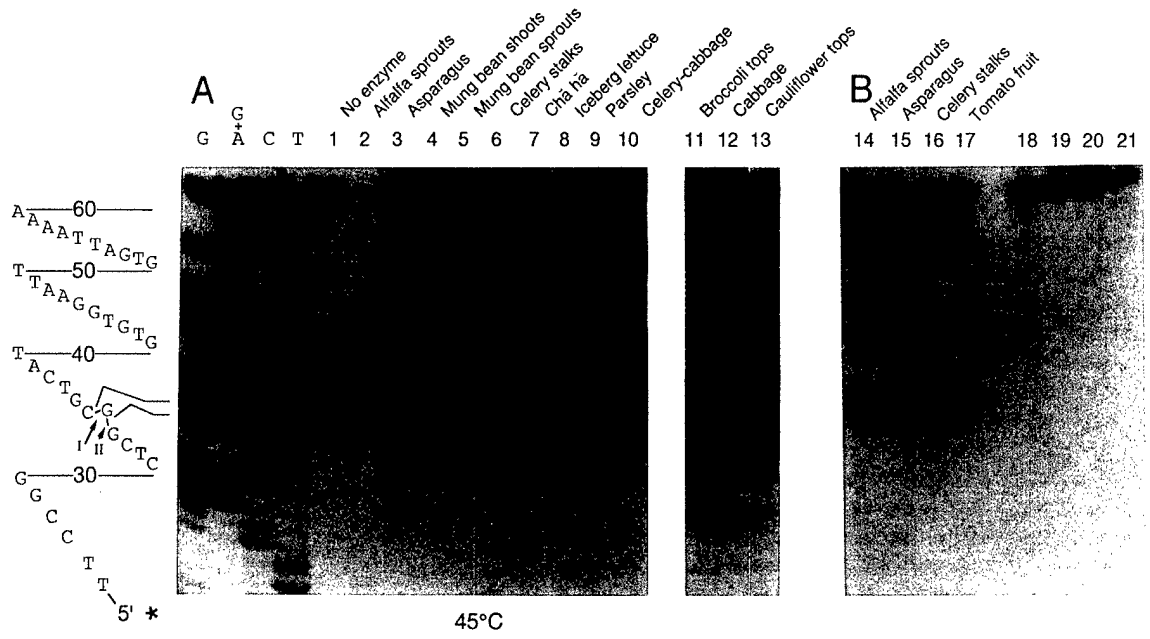


Figure 2. Conserved features of the CEL I-like mismatch endonucleases from different plants. (**A**) One microliter of plant extract was used in each incubation with a mismatch duplex containing an extrahelical G residue. The substrate was 5′-labeled in the top strand and incubation was at 45°C. (**B**) One milliliter of each of the crude extracts of the plants was applied to a 100 µl column of concanavalin A–Sepharose resin (Sigma) in 20 mM HEPES, pH 7.0, 0.5 M KCl buffer, washed and eluted with 200 µl 0.5 M α-D+-mannose in 0.5 M KCl, pH 7.0. One microliter of the eluted enzyme was used in the reactions in lanes 14–21. Lanes 18–21 were control reactions for lanes 14–17, respectively, using the perfectly base paired substrate.

RESULTS

Detection of CEL I-like activities in plant extracts

By incubating plant extracts with a mismatch-containing heteroduplex, we detected a novel mismatch endonuclease activity. This activity performs a single-strand cut on the 3′-side of a mismatch site (Fig. 2). The activity appears to be present in many common vegetables and in a variety of plant tissues: root, stem, leaf, flower and fruit. From each tissue, we have found a similar amount of mismatch endonuclease activity per gram of tissue (Fig. 2A, lanes 2–13). We named the prominent activity present in celery CEL I. The substrate initially used was a 5′-labeled duplex with an extrahelical G nucleotide mismatch that can alternate between two consecutive G residues, thereby giving two CEL I cut bands. These gel mobilities are consistent with the production of a 3′-OH group on the deoxyribose moiety (6). All the CEL I-like mismatch endonucleases cut the DNA at the same two alternate positions on the 3′-side of the mismatch. The mismatch endonucleases of alfalfa sprout, asparagus, celery and tomato were each found to bind to a concanavalin A–agarose column and were eluted by α-D+-mannose (Fig. 2B). Thus, CEL I-like activities appear to be mannosyl glycoproteins.

Purification of CEL I

Celery stalks were chosen to be a source of model enzyme because of the year-round availability of celery, a low amount of chloroplast proteins and pigments in the extracts and the high mismatch specificity of CEL I. The CEL I purification procedure started with celery juice, containing ~350 g protein, from 7 kg celery stalk. The Superose 12 fraction contained 3 ml CEL I at 0.1 µg/µl and is estimated to be ~10 000-fold purified with a recovery of 9%. SDS–PAGE followed by staining with Coomassie Blue R250 indicated that the purest CEL I contains more than one protein band of 34–39 kDa (data not shown). It is not clear yet whether these bands represent glycoforms of CEL I or whether proteins with unrelated properties are present.

Incision by CEL I at mismatches of single nucleotide DNA loops and nucleotide substitutions

The mismatch incision by purified CEL I in substrates containing a single extrahelical nucleotide is shown in Figure 3A (lanes 2–5). This analysis shows that CEL I has a preference for G > A > C > T in the extrahelical position. The activity of CEL I is stimulated by the presence of Taq DNA polymerase (Fig. 3A, lanes 6–10). This stimulation of CEL I does not require dNTP (data not shown). Taq DNA polymerase stimulation of incision at the weak extrahelical T substrate is ~30-fold (Fig. 3A, comparing lanes 5 and 10), as measured by densitometry of the autoradiogram bands (data not shown). The DNA polymerase stimulation is less for extrahelical G and A substrates (Fig. 3A, lanes 7 and 8, respectively) because these substrates are already efficiently cut by CEL I. Because of base pairing slippage in the extrahelical nucleotide G and C substrates (Fig. 3A, lanes 2 and 4), two incision bands were seen. At the extrahelical nucleotide that is closer to the 5′-terminus, in the presence of Taq DNA polymerase and dNTP in lanes 7 and 9 mismatch slippage allows nick translation to occur after CEL I incision. As a result, the lower

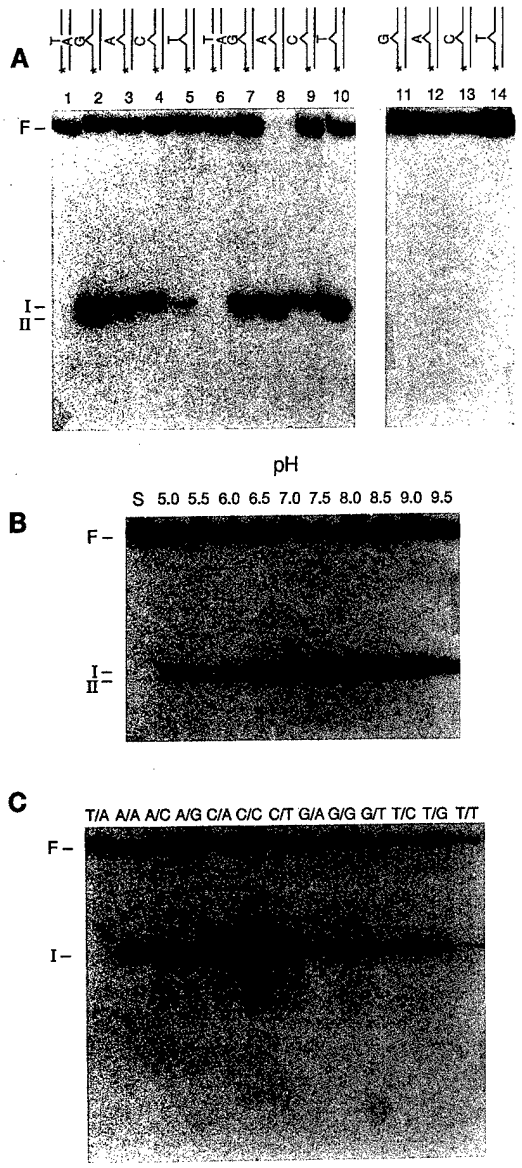


Figure 3. Mismatch incision of the purified CEL I nuclease at different mismatches. (**A**) Taq DNA polymerase stimulation of purified CEL I incision at DNA mismatches of a single extrahelical nucleotide. Autoradiograms of denaturing 15% polyacrylamide gels are shown. F, full-length substrate, 65 nt long, labeled at the 5′-terminus (*) of the top strand. Lanes 1–5 and 6–10, 50 fmol substrates, in the presence of 10 μM dNTP, treated with 20 ng purified CEL I, without and with 0.5 U Taq DNA polymerase, respectively, for 30 min at 37°C; lanes 1 and 6, substrates containing no mismatch; lanes 11–14, substrates incubated with only Taq DNA polymerase in the presence of 10 μM dNTP, with the autoradiogram exposure time extended 3×. The two cuts (I and II) in lanes 2 and 4 are due to mismatch slippage in alternative base pairing possibilities. One mismatched base at each cut site was repaired by DNA polymerase + dNTP in lanes 7 and 9. (**B**) pH profile of CEL I mismatch incision at a substrate with a single extrahelical G residue. S, substrate incubated without CEL I. Taq DNA polymerase and dNTP were not present in this study. If Taq DNA polymerase, but not dNTP, were included, the pH profile is similar, but the incision efficiency would be near completion in all lanes (data not shown). (**C**) CEL I incision at base substitutions. The top strands were 5′-labeled. Incubation with CEL I was for 30 min at 45°C in the presence of Taq DNA polymerase but no dNTP.

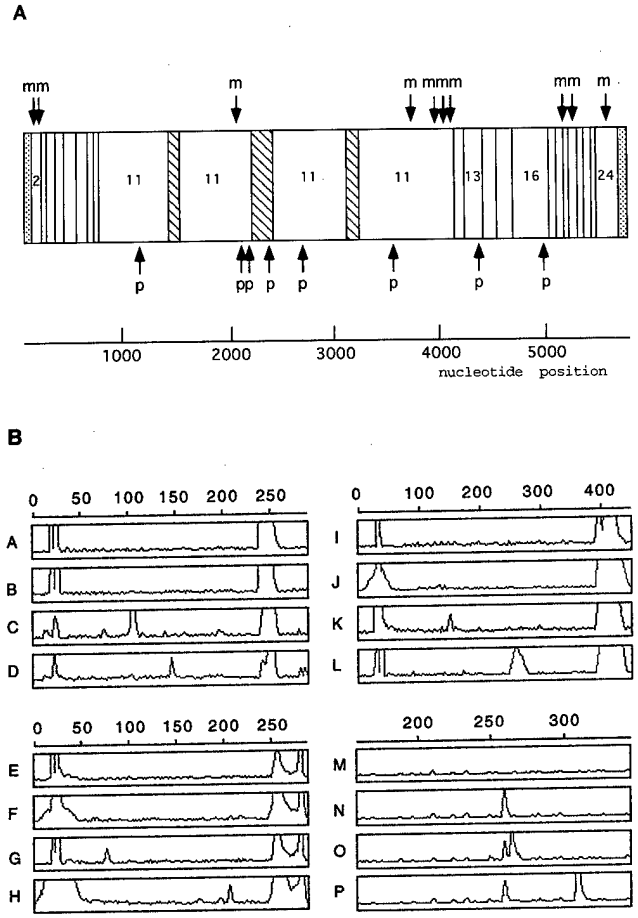


Figure 4. CEL I enzymatic mutation detection in the *BRCA1* gene. (**A**) Schematic presentation of the exons of the *BRCA1* gene and the polymorphisms and mutations detected in this report. The *BRCA1* gene is divided into 24 exons (22 coding exons). For CEL I mutation detection, each PCR usually covers one exon. Exon 11 is divided into four regions of ~1000 bp that overlap by at least 100 bp indicated by the diagonally shaded areas. All of exon 1, part of exon 2 and part of exon 24 are untranslated regions, as denoted by dotted areas. Exon 4 is not part of the mRNA (7). p, polymorphisms; m, mutations. (**B**) Electropherogram of CEL I mutation detection GeneScan analyses. Two color fluorescent heteroduplexes of the PCR products of the *BRCA1* gene were prepared as described in Materials and Methods. All lanes have CEL I treatment. Vertical axis, relative fluorescence units; horizontal axis, DNA length in nucleotides. (A–D) Deletion of A in exon 19. The CEL I mismatch-specific peaks seen at sizes 106 and 146 nt in (C) and (D) for the 6-FAM-labeled and the TET-labeled strand, respectively, were not present in the wild-type control for the FAM (A) and the TET (B) strands. Full-length PCR product was observed at 249 nt length and residual primers at 20–30 nt. The signal in the full-length position exceeded the linear range of the detector. (E–H) Detection of C→T base substitution in exon 24. The PCR product was 286 bp. This C→T base substitution was detected as blue at fragment sizes 76 and 77 nt in (G) and as green at fragment size 206 nt in (H) for the 6-FAM-labeled and the TET-labeled strand, respectively, but not in the wild-type control for the FAM (E) and the TET (F) strands. (I–L) Detection of a C insertion mutation in exon 20. The PCR product was 410 bp long. This insertion of a single C residue was detected at fragment sizes 151 and 259 nt for exon 20 in (K) and (L), respectively, for the 6-FAM-labeled and the TET-labeled strand, respectively. The mutation-specific CEL I cuts were not observed in the wild-type controls for the FAM (I) and the TET (J) strands. (M–P) Detection of mutations next to a polymorphism in exon 11. The PCR product was 1006 bp long. (M) Wild-type control treated with CEL I. (N) Polymorphism (2201 T→C) identified by CEL I. (O) Two polymorphisms (2210 T→C and 2196 G→A) detected by CEL I. (P) Polymorphism (2210 T→C) and mutation (K630ter; 2154A→T) detected by CEL I. Only the data from the TET-labeled strands are presented in (M–P).

nt #	Exon #	p/m	DNA change	5′ sequence	Heteroduplex formation		3′ sequence
185	2	m	AG deletion	5′ ATCTT	AG	TC	5′ AGTGT
188	2	m	11 bp deletion	5′ TTAGA	GTGTCCCATCT	CACAGGGTAGA	5′ GGTAA
1186	11	p	A → G	5′ TAAGC	A/C	G/T	5′ GAAAC
2154	11	m	A → T	5′ GAGCC	A/A	T/T	5′ AGAAG
2196	11	p	G → A	5′ GACAT	G/T	A/C	5′ ACAGC
2201	11	p	T → C	5′ GACAG	T/G	C/A	5′ GATAC
2430	11	p	T → C	5′ AGTAG	T/G	C/A	5′ AGTAT
2731	11	p	C → T	5′ TGCTC	C/A	T/G	5′ GTTTT
3667	11	p	A → G	5′ CAGAA	A/C	G/T	5′ GGAGA
3819	11	m	5 bp deletion	5′ GTAAA	GTAAA	CATTT	5′ CAATA
4153	11	m	A deletion	5′ TGATG	A	T	5′ AGAAA
4168	11	m	A → G	5′ AACGG	A/C	G/T	5′ CTTGA
4184	11	m	4 bp deletion	5′ AATAA	TCAA	AGTT	5′ GAAGA
4427	13	p	T → C	5′ GACTC	T/G	C/A	5′ TCTGC
4956	16	p	A → G	5′ CCCAG	A/C	G/T	5′ GTCCA
Intron 18	19	m	A deletion	5′ TCTTT	A	T	5′ GGGGT
5382	20	m	C insertion	5′ ATCCC	C	A	5′ AGGAC
5622	24	m	C → T	5′ TGACC	C/A	T/G	5′ GAGAG

Figure 5. Summary of mutations and polymorphisms detected in the *BRCA1* gene by CEL I in this study. m, mutation; p, polymorphism.

band of CEL I incision seen in lanes 2 and 4 was restored to full-length in lanes 7 and 9.

pH optimum of CEL I endonuclease

The pH optimum of CEL I appears to be in the neutral range although the enzyme is active from pH 5 to 9.5 The pH activity profile of CEL I cutting of the extrahelical G mismatch substrate without Taq DNA polymerase stimulation is shown in Figure 3B.

Incisions of CEL I at base substitutions

Base substitution mismatched substrates are also recognized by CEL I and cut on one of the two DNA strands for each mismatch duplex (Fig. 3C). Some of these substrates are less efficiently incised compared with those containing DNA loops. For the purpose of mutation detection *in vivo*, all base substitution mismatches can be detected by CEL I at 45°C in the presence of 0.5 U Taq DNA polymerase (Fig. 3C). Substrates with the 5′-terminus of the top strands labeled were used in this study. CEL I substrate preference shown here is C/C ≥ C/A ~ C/T ≥ G/G > A/C ~ A/A ~ T/C > T/G ~ G/T ~ G/A ~ A/G > T/T.

Detection of mutations and polymorphisms in the *BRCA1* gene

A CEL I-based assay was used to detect mutations and polymorphisms in various exons of the *BRCA1* gene (Fig. 4). Strong incision bands were observed for heteroduplex alleles but not for wild-type alleles (Fig. 4B). The CEL I assay is also capable of detecting multiple sequence variants within the same DNA strand (Fig. 4, panels M–P).

A summary of the mutations and polymorphisms in the *BRCA1* gene detected by CEL I in this study is shown in Figure 5. Sequence analyses of the coding regions and intron/exon boundaries confirmed that all known sequence variants were detected by CEL I. The DNA sequences flanking each mutation or polymorphism illustrate that CEL I detects mismatches in a variety of sequence contexts. Furthermore, no false positive or false negative conclusions were encountered.

DISCUSSION

Plants and fungi contain single-stranded specific nucleases that attack both DNA and RNA (8). S1 nuclease from *Aspergillus oryzae* (1), P1 nuclease from *Penicillium citrinum* (9) and mung bean nuclease from the sprouts of *Vigna radiata* (2–3) are Zn proteins active mainly near pH 5.0. CEL I is similar to these enzymes in that the most purified enzyme fraction shows some single-stranded DNase activity and endonuclease activity on supercoiled plasmids, relaxed double-stranded DNA, UV irradiated plasmids and Y-shaped DNA duplexes (data not shown). However, CEL I is most active on mismatch substrates. The neutral pH optimum, incision primarily at the phosphodiester bond immediately on the 3′-side of the mismatch and stimulation of activity by a DNA polymerase are properties that distinguish CEL I from the above nucleases. The mechanism for DNA polymerase stimulation of the CEL I activity is presently unknown. One possibility is that DNA polymerase has a high affinity for the 3′-OH group produced by CEL I incision at the mismatch and displaces CEL I simply by competition for the site. Such protein displacement will allow CEL I to recycle catalytically. For the purpose of mutation detection, DNA polymerases with 3′→5′ exonuclease proofreading activity cannot be used. Such

DNA polymerases, of which the Klenow fragment of *E.coli* DNA polymerase I is an example, will excise the mismatch nucleotide after DNA polymerase displacement of CEL I at the site of mismatch incision. In the absence of dNTP, one will observe 3′→5′ exonuclease degradation of the DNA fragment produced by CEL I mismatch incision. In the presence of dNTP, a highly efficient *in vitro* mismatch correction system will have been reconstituted (data not shown). It is necessary to test whether or not other proteins, such as DNA helicases, DNA ligases and DNA terminus-binding proteins, can also assist CEL I at mismatch incision *in vivo*.

In the CEL I detection scheme used in this paper, two alleles will form two alternate heteroduplex mispairs such that at least one mismatch in each pair should be a good substrate for CEL I. G/G is paired with C/C, A/G is paired with C/T, A/C is paired with G/T and T/T is recognized least well by CEL I, but an A/A mismatch will be present in such a heteroduplex preparation and will be detected by CEL I. As shown in Figure 5, flanking sequence context apparently does not adversely affect the ability of CEL I to identify a mutation. Even mismatches flanked by GC-rich regions (Fig. 1) are recognized. The four PCR products of *BRCA1* exon 11 are 889–1120 bp in length. Most of the time, mismatch incision will be observed as both colors in the electropherogram such that each independently confirms the position of the mutation/polymorphism. The sum of the two fragments is theoretically 1 nt more than the length of the PCR product. In the cases of mismatches that can wobble in alternative base pairings because of the sequence contexts and for large DNA loops the sum of the two fragments may deviate from the above rule.

The principle of mismatch recognition by CEL I appears to be different from T_4 endonuclease VII, which has also been used for enzyme mutation detection (11,12). The latter is a resolvase, which nicks one strand at the site of a mismatch and then in the other strand across from the DNA nick (12). Therefore, any nick can produce two corresponding fragments of the two colors. In the case of CEL I, the two fragments of the two colors represent two truly independent mutation detection events that complement each other to confirm the presence of the mutation. This distinction is because CEL I only nicks one strand of DNA in a mismatch heteroduplex at the site of the mismatch. There is no second cut in the opposite strand of the same DNA molecule after the first nick. Moreover, the CEL I mechanism allows the non-cut strand to be potentially useful as template for the removal of non-specific nicks, if any, by nick translation repair or ligation. Unlike resolvases, CEL I shows no tendency to nick duplex DNA at unique DNA sequences.

Other strengths of the CEL I mutation detection assay are its simplicity and its lack of preference for unique non-mismatch DNA sequences. Background non-specific DNA nicking is very low. The high signal-to-noise ratio of CEL I using fluorescent dye-labeled PCR products often allows mutations to be detected by visual inspection of the GeneScan gel image. CEL I is a very stable enzyme, during both its purification, storage and assay.

CEL I mutation detection provides a mutation detection method based on different principles than DNA sequencing and single-strand conformation polymorphism (SSCP) (13). In genes such as *BRCA1*, mutations can occur in numerous positions, making it very difficult for most mutation detection methods to screen for mutations in this gene. To date, >520 individual sequence alterations are known in the *BRCA1* gene. The ability of CEL I to detect a mismatch at any one or more nucleotide positions without prior knowledge of the mutation provides promise of a very powerful method for screening mutations in cancer genes. Indeed, the ease of setting up and performing CEL I mutation detection should allow it to be established quickly in most laboratories.

ACKNOWLEDGEMENTS

This work was supported, in part, by grants NIH CA71426 and US Army DMAD17-97-1-7286 to A.T.Y., NIH CA70328 and US ARMY DAMD17-96-1-6088 to A.K.G., by institutional grants from the National Institutes of Health to the Fox Chase Cancer Center (CA06927, RR05539), an appropriation from the Commonwealth of Pennsylvania and a grant from the Glenmede Trust to the FCCC. The authors thank Dr Alfonso Bellacosa for critical reading of this manuscript.

REFERENCES

1 Shank,T.E., Rhodes,C., Rigby,P.W.J. and Berg,P. (1975) *Proc. Natl Acad. Sci. USA*, **72**, 989–993.
2 Kowalski,D., Kroeker,W.D. and Laskowski,M.,Sr (1976) *Biochemistry*, **15**, 4457–4462.
3 Sheflin,L.G. and Kowalski,D. (1984) *Nucleic Acids Res.*, **12**, 7087–7104.
4 Loeb,L.A. and Silber,J.R. (1981) *Biochim. Biophys. Acta*, **656**, 256–264.
5 Kowalski,D. and Sanford,J.P. (1982) *J. Biol. Chem.*, **257**, 7820–7825.
6 Jones,B.K. and Yeung,A.T. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 8410–8414.
7 Miki,Y., Swenson,J., Shattuck-Eidens,D., Futreal,P.A., Harshman,K., Tavtigian,S., Liu,Q., Cochran,C., Nennett,L.M., Ding,W., Bell,R., Rosenthal,J., Hussey,C., Tran,T., McClure,M., Frye,C., Hattier,T., Phelps,R., Haugen-Strano,A., Katcher,H., Yakumo,K., Gholami,Z., Shaffer,D., Stone,S., Bayer,S., Wray,C., Bogden,R., Dayananth,P., Ward,J., Tonnin,P., Narod,S., Bristow,P.K., Norris,F.H., Helvering,L., Morrison,P., Rosteck,P., Lai,M., Barrett,J.C., Lewis,C., Neuhausen,S., Cannon-Albright,L., Golgar,D., Wiseman,R., Kamb,A. and Skolnick,M.H. (1994) *Science*, **266**, 66–71.
8 Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds) (1993) *Nucleases*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9 Maekawa,K., Tsunasawa,S., Dibo,G. and Sakiyama,F. (1991) *Eur. J. Biochem.*, **200**, 651–661.
10 Youil,R., Kemper,B. and Cotton,R.G.H. (1996) *Genomics*, **32**, 431–435.
11 Del Tito,B.J.,Jr, Poff,H.E., Novotny,M.A., Cartledge,D.M., Walkwer,R.I., Earl,C.D. and Bailey,A.L. (1998) *Clin. Chem.*, **44**, 731–739.
12 Solaro,P., Birkenkamp,K., Pfeiffer,P. and Kemper,B. (1993) *J. Mol. Biol.*, **230**, 868–877.
13 Orita,M., Iwahana,H., Kanazawa,H., Hayashi,K. and Sekiya,T. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2766–2770.

Incision at Nucleotide Insertions/Deletions and Base Pair Mismatches by the SP Nuclease of Spinach

Catherine A. Oleykowski, Colleen R. Bronson Mullins, David W. Chang, and Anthony T. Yeung

Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Biochemistry®

Reprinted from
Volume 38, Number 7, Pages 2200–2205

Incision at Nucleotide Insertions/Deletions and Base Pair Mismatches by the SP Nuclease of Spinach†

Catherine A. Oleykowski, Colleen R. Bronson Mullins, David W. Chang,‡ and Anthony T. Yeung*

Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Received September 25, 1998; Revised Manuscript Received December 8, 1998

ABSTRACT: Spinach leaves contain a highly active nuclease called SP. The purified enzyme incises single-stranded DNA, RNA, and double-stranded DNA that has been destabilized by A-T-rich regions and DNA lesions [Strickland et al. (1991) *Biochemistry 30*, 9749−9756]. This broad range of activity has suggested that SP may be similar to a family of nucleases represented by S1, P1, and the mung bean nuclease. However, unlike these single-stranded nucleases that require acidic pH and low ionic strength conditions, SP has a neutral pH optimum and is active over a wide range of salt concentrations. We have extended these findings and showed that an outstanding substrate for SP is a mismatched DNA duplex. For base-substitution mismatches, SP incises at all mismatches except those containing a guanine residue. SP also cuts at insertion/deletions of one or more nucleotides. Where the extrahelical DNA loop contains one nucleotide, the preference of extrahelical nucleotide is A ≫ T ∼ C but undetectable at G. The inability of SP to cut at guanine residues and the favoring of A-T-rich regions distinguish SP from the CEL I family of neutral pH mismatch endonucleases recently discovered in celery and other plants [Oleykowski et al. (1998) *Nucleic Acids Res. 26*, 4597−4602]. SP, like CEL I, does not turn over after incision at a mismatched site in vitro. Similar to CEL I, the presence of a DNA polymerase or a DNA ligase allows SP to turn over and stimulate its activity in vitro by about 20-fold. The possibility that the SP nuclease may be a natural variant of the CEL I family of mismatch endonucleases is discussed.

Nucleases participate in many essential cellular functions (*1*). Some nucleases are highly specialized in DNA recombination and repair while others enable general degradation of dietary nucleic acids. Of the latter, the secreted fungal nucleases, S1 (*2*) and P1 (*3*), and the pancreatic DNase I (*4*) are the best characterized. Often, a nuclease may possess multiple activities within one polypeptide, thus enabling it to perform both general nucleic acid degradation and unique steps in DNA replication, recombination, or repair. For example, Exo III of *E. coli* is a powerful 3′ to 5′ exonuclease as well as being the major apurinic endonuclease in this organism and a 3′ phosphatase (*5*). The recBCD recombination nuclease is a potent 5′ to 3′ and 3′ to 5′ exonuclease and a helicase (*6*, *7*).

Spinach (*Spinacia oleracea*) contains a nuclease called SP (*11*, *12*) that has multiple activities. The purified SP, similar to S1, P1, and mung bean nuclease (*13*−*15*), is able to degrade single-stranded DNA, double-stranded DNA, and RNA. Instead of having an acidic pH optimum like S1, P1, and mung bean nuclease, SP has a neutral pH optimum. Interestingly, SP incises DNA-containing cisplatin adducts, the $TC_{(6-4)}$-type pyrimidine dimers, but not the cyclobutane-type pyrimidine dimers (*12*). Such properties suggest that SP could be a repair enzyme. We report here an unexpected prominent property of SP: the incision at DNA insertion/deletion loops, and at base-substitution mismatches, under physiological conditions. We also show that in vitro SP mismatch incision activity is stimulated by the presence of a DNA polymerase or a DNA ligase.

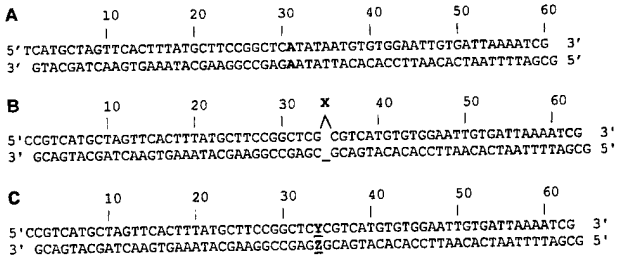


FIGURE 1: Heteroduplex DNA substrates. (A) The oligonucleotide duplex substrate containing an A/A mismatch at position 30, next to an A-T-rich region. (B) The substrates containing an extrahelical DNA loop X with one or more nucleotides. (C) The base-substitution substrates. Y and Z are various nucleotides that can be substituted in to produce the mismatches used in this study.

EXPERIMENTAL PROCEDURES

SP Nuclease. SP nuclease for the initial mismatch endonuclease assays was generously provided by Dr. Doetsch of Emory University. Subsequent experiments used SP prepared in our laboratory according to the published protocol (*11*).

Preparation of Plant Extracts. Various plant tissues were homogenized in a blender at 4 °C and adjusted with a 10× solution to give the composition of buffer A [0.1 M Tris-

† This work was supported, in part, by NIH Grant CA71426 and U.S. Army Grant DMAD17-97-1-7286 to A.T.Y., by institutional grants from the National Institutes of Health to Fox Chase Cancer Center (CA06927, RR05539), and by an appropriation from the Commonwealth of Pennsylvania.

* To whom correspondence should be addressed. Telephone: (215) 728-2488. FAX: (215) 728-3647. Email: AT_Yeung@fccc.edu.

‡ Current address: Princeton University.

10.1021/bi982318y CCC: $18.00
Published on Web 01/26/1999

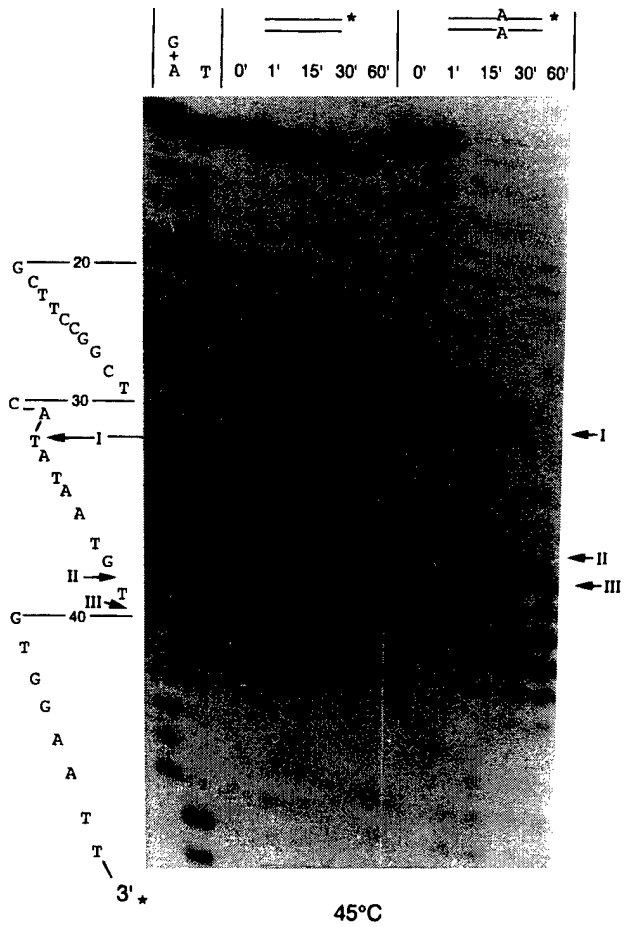


FIGURE 2: Incision of mismatched substrates by the spinach SP nuclease. Fifty femtomoles of no-mismatch substrate, or A/A base-substitution substrate of Figure 1A, 3′-labeled (*) in the top strand was incubated with 1 ng of SP in buffer C for various durations at 45 °C. The 20 μL reaction was terminated by adding 10 μL of 1.5% SDS, 47 mM EDTA, and 75% formamide plus tracking dyes and analyzed on a denaturing polyacrylamide gel in 7 M urea and 50 °C. The autoradiogram is shown. Chemical DNA sequencing ladders are shown for determining the positions of incisions (I, II, and III) in the DNA sequence. For a 3′-labeled substrate, when a nuclease nicks 3′ of a nucleotide and produces a 5′-PO_4 terminus, the labeled truncated band comigrates with the band for that nucleotide in the chemical DNA sequencing reaction lane (*13*).

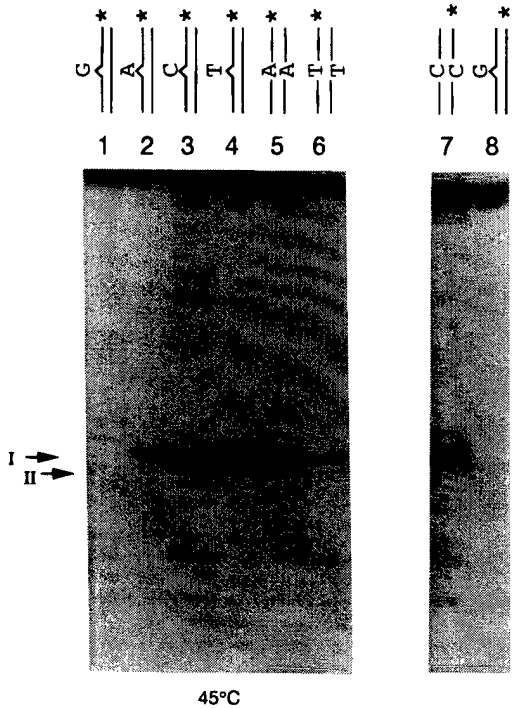


FIGURE 3: SP incision at various mismatches. Substrates were made from the sequences in Figure 1B,C with the position of the radioactivity indicated by an asterisk. Fifty femtomoles of DNA substrate was incubated with 5 ng of SP in buffer B for 30 min at 45 °C. Lanes 1−6 were 3′-labeled in the top strand. Lanes 7 and 8 were 5′-labeled in the bottom strand. SP incision produced bands I and II about 32 nucleotides (nt) long. The substrates with the extrahelical nucleotides are shown in lanes 1−4. Base-substitution heteroduplexes are shown in lanes 5−7. I = incision position at the 3′ side of the extrahelical A nucleotide. II = the incision from an alternate extrahelical C base-pairing permissible in this sequence.

HCl, pH 7.7, plus 10 μM phenylmethanesulfonyl fluoride (PMSF)]. The extracts were stored at −70 °C. Alternatively, the tissues were frozen in liquid nitrogen, ground to a powder with a mortar and pestle, and then extracted with buffer A on ice. Both types of extract provided equivalent data.

Preparation of Extracts of Spinach Seedlings. Seeds of several spinach varieties were purchased from gardening centers in Philadelphia, PA. The seeds were soaked for 3 h in water and planted in soilless potting soil and allowed to grow for 3 weeks before harvest. The plant tissues were frozen in liquid nitrogen and ground to a powder in a liquid nitrogen-cooled mortar and pestle. The powder was extracted with buffer A and stored at −70 °C.

Preparation of Heteroduplexes Containing Various Mismatches. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and purified by electrophoresis in a denaturing polyacrylamide gel in the presence of 7 M urea at 50 °C. The purified single-strand oligonucleotides were hybridized with appropriate opposite strands to construct DNA heteroduplex substrates, 61−65 bp[1] long containing base-substitution mismatches, or insertion/deletion DNA loops (Figure 1). The DNA duplexes were labeled with ^{32}P at one of the four termini so that DNA endonuclease incisions at the mispaired nucleotides could be identified as truncated DNA bands on denaturing DNA sequencing gels (16). The 5′-labeled substrates were labeled as single-strand DNA with T4 polynucleotide kinase and [γ-^{32}P]ATP before annealing to its opposite strand. The 3′-labeled substrates were labeled by the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP and [α-^{32}P]dGTP after annealing. All the labeled duplexes were made blunt-ended by the fill-in reaction of DNA polymerase I Klenow fragment using dCTP and dGTP, and purified by electrophoresis in a nondenaturing polyacrylamide gel as described (16). DNA was electroeluted from the gel slice in a Centricon unit with an AMICON Model 57005 electroeluter. The upper reservoir of this unit has been replaced with one having watertight partitions to prevent cross-contamination.

Mismatch Endonuclease Assay. Ten to fifty femtomoles of ^{32}P-labeled substrates was incubated with 0.3−5 ng of the purified SP preparation in buffer B (20 mM Tris-HCl, pH 7.4, 25 mM KCl, 10 mM $MgCl_2$) for 30 min. The 20 μL reaction was terminated by adding 10 μL of 1.5% SDS, 47 mM EDTA, and 75% formamide plus tracking dyes and analyzed by denaturing polyacrylamide gel electrophoresis

[1] Abbreviations: bp, base pair; nt, nucleotide(s); PCR, polymerase chain reaction.

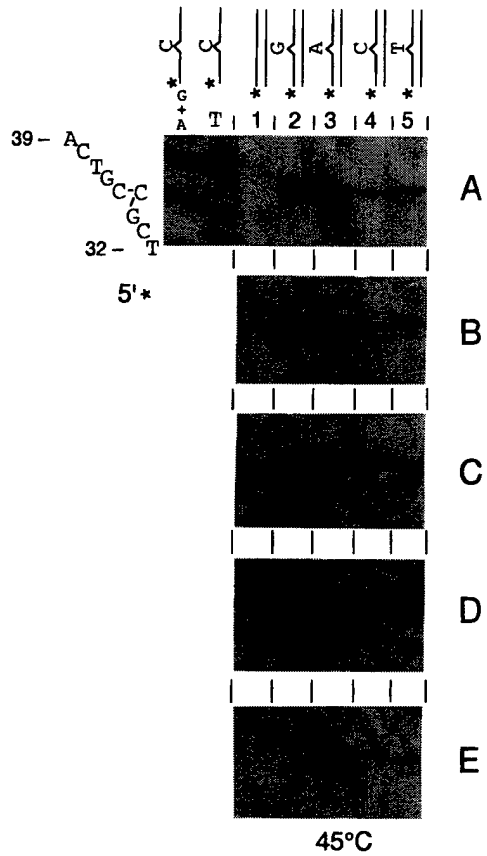


FIGURE 4: Comparison of various plant extracts for the ability to incise at extrahelical nucleotide residues. Fifty femtomoles of each substrate labeled at the 5′ termini of the top strand was incubated with 1 μL of various plant extracts for 30 min at 45 °C. The treated DNA was analyzed as described the experiment in Figure 2. Panels A, B, C, D, and E represent extracts from broccoli, cabbage, cauliflower, celery, and spinach, respectively. Chemical DNA sequencing ladders are used for deducing the positions of the incisions in the substrates. For a 5′-labeled substrate, when a nuclease nicks 5′ of a nucleotide and produces a 3′-OH terminus, the truncated band migrates half nucleotide spacing slower than the band for that nucleotide in the lane containing the products of the chemical DNA sequencing reaction (*13*).

as described (*13*). When Ampligase (a thermostable DNA ligase from Epicentre Technologies) was present in the assay, the reaction was carried out in buffer C (buffer B plus 0.001% Triton X-100 and 6 mM NAD).

RESULTS AND DISCUSSION

Incision of SP at an A/A Mismatch Located Next to an A-T-Rich Region. Incubation of SP with an A/A mismatch-containing substrate results in incision near the mismatch (Figure 2). SP is known to cleave DNA to produce 3′-hydroxyl and 5′-phosphoryl termini (*12*). The incisions at 1 min in the 3′-labeled top stand were traced to the first and second phosphodiester bonds 3′ of the A/A mismatch site. The shorter bands at later time points were probably produced by further SP digestion of the A-T-rich region destabilized by the DNA nicking at the mismatch.

Low-level nonspecific DNA nicking of the nonmismatched substrate by SP at the A-T-rich region destabilized by 45 °C (Figure 2) shows that SP exhibits properties similar to those of S1 type of single-strand-specific nucleases, except it does so at neutral pH.

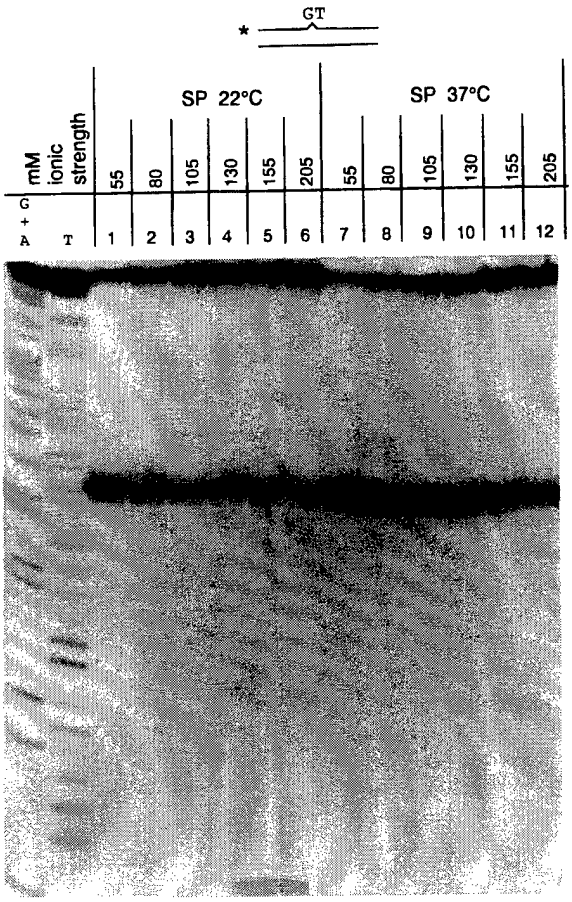
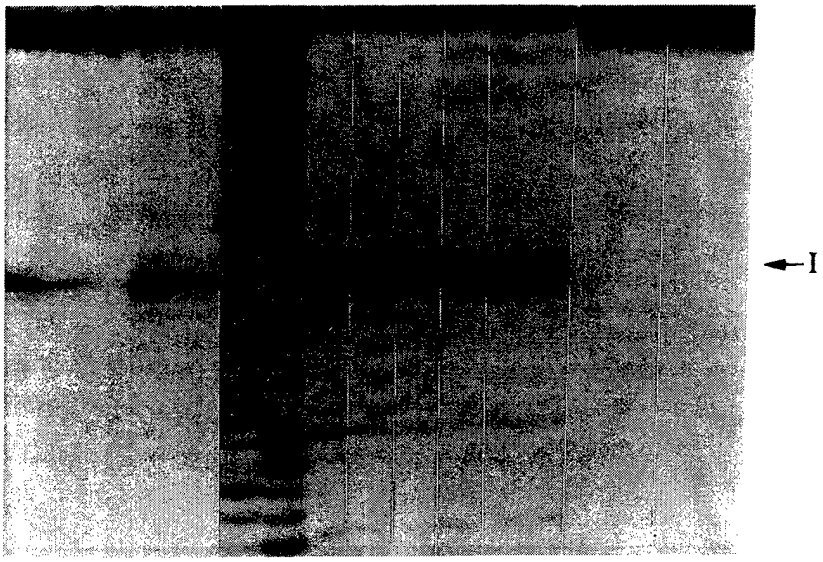


FIGURE 5: Ability of SP to incise at a GT loop mismatch under physiological conditions. Fifty femtomoles of 5′ top-strand-labeled extrahelical GT loop substrate was incubated with 5 ng of SP for 30 min at 22 or 37 °C. NaCl was added to buffer B to reach various ionic strengths. Procedures were as described in Figure 2.

Because A-T-rich regions enhance the ability of incision at a mismatch, we redesigned the substrates to include the challenge of G-C-rich flanking sequences (Figure 1B,C). These G-C-rich substrates are used in all experiments in Figures 3−7. DNA nibbling from the mismatch cut site does not occur for these substrates in which the mismatch is not in an A-T-rich region. Figure 3 illustrates the result of SP incision at substrates containing an extrahelical nucleotide of A, C, or T residue. Incision was not observed in the substrate with an extrahelical G residue. The absence of cutting in the top strand of this substrate is not due to the possibility of cutting being directed to the bottom strand. In fact, incision was not observed in the bottom strand of this substrate either (lane 8). In another control experiment to evaluate the possibility that flanking G-C-rich sequences may have inhibited SP cutting at guanine residues, we found that SP did not incise at either a single guanine residue or a loop of five guanine residues, inserted in an A-T-rich region (data not shown). Thus, the reactivity of SP with insertion/deletion mismatches is consistent with the known intrinsic preference of SP for A and T residues (*12*).

In lane 3 of Figure 3, the extrahelical C substrate produced a band one nucleotide shorter than the extrahelical A and T substrates in lanes 2 and 4, respectively. The likely reason is that the extrahelical C in this substrate was located 5′ to another C residue, therefore allowing the two C residues to alternate in base-pairing with the G residue in the opposite

SP	4x	2x	1x	4x	2x			4x	2x	1x	1x	2x	1x	–	–	–	–
Taq	–	–	–	–	–			4x	4x	4x	1x	2x	1x	4x	2x	4x	2x
Lig	–	–	–	+	+	C	T	–	–	–	–	+	+	–	–	+	+



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

FIGURE 6: Stimulation of SP by DNA polymerase and DNA ligase. Lanes 1−3 are a dilution series of SP, 1.2, 0.6, and 0.3 ng, respectively, incubated with 50 fmol of top strand 5′-labeled A/A-mismatched heteroduplex for 30 min at 45 °C. SP nicking of the mismatch was barely visible at 0.3 ng. However, if 12.5 units of Ampligase were present, and/or 0.125 unit (1×) of Taq DNA polymerase was present, the SP cutting at the mismatch was greatly stimulated. I = incision site at the first phosphodiester bond 3′ of a mismatch. C and T chemical DNA sequencing ladders are shown for position reference.

strand. One of these C mismatch conformations is favored in the reaction.

Lanes 5, 6, and 7 of Figure 3 showed that A/A, T/T, and C/C mismatches are also incised, but T/T is cut less well by SP. Other base-substitution mismatches were tested under these conditions or under more favorable conditions and will be described after we show how the favorable conditions were established.

To ascertain whether the lack of incision at guanine nucleotides by the purified SP was due to the loss of some activity or protein factors during enzyme purification, we performed an assay of the crude cell extract (Figure 4). Extracts of broccoli, cabbage, cauliflower, and celery were able to incise at all four substrates containing an extrahelical nucleotide (panels A−D, lanes 2−5) without significant background cutting in the no-mismatch substrate (lane 1), as expected due to the presence of a CEL I-like activity (*16*). In contrast, spinach extract (panel E) failed to incise at the substrate containing the extrahelical G residue, but it was able to cut the three substrates with A, C, or T extrahelical nucleotides. As a further survey, extracts of 3-week-old seedlings of six varieties of spinach were tested with this extrahelical G substrate, and all were found to be unable to incise at this mismatch (data not shown). The spinach varieties tested were all of 1998 lots: Avon Hybrid, Melody Hybrid, TYEE, Indian Summer Hybrid, and two varieties of Bloomsdale Long-Standing spinach. This finding suggests that the inability of SP to cut at a G mismatch is not unique to one variety of spinach.

Incision of Mismatch by SP under Physiological Conditions. Figure 5 shows that SP is efficient at mismatch recognition at 22 and 37 °C under a variety of ionic strength conditions and neutral pH. These conditions are not known to favor the mechanisms of S1, P1, and mung bean nuclease type nucleases. For example, S1 nuclease does not cleave DNA at single-base mismatches at pH 4.6 (17) or pH 7.5 (data not shown). The efficiency of mung bean nuclease to nick supercoiled DNA is 5 orders of magnitude higher at pH 5 than at pH 8 (18). In this substrate containing two extrahelical GT nucleotides, the incision by SP occurred between the GT dinucleotides, at the 3′ side of the G residue, as determined by comparison with the chemical DNA sequencing ladder on the side (19). Whether the apparent incision position 3′ of a G residue is the result of incision at the 3′ side of the T residue, followed by exonuclease removal of the T residue, was not tested. It will be interesting to elucidate the parameters that govern the nucleotide specificity of these nucleases.

Mechanism of Turnover of the SP Nuclease. Figure 6 illustrates the ability of Taq DNA polymerase and Ampligase to stimulate SP activity. In Figure 6, lanes 1−3, decreasing amounts of SP were incubated with 50 fmol of A/A-mismatched substrate for 45 min. The SP mismatch-specific incision band was barely visible in lane 3. In lanes 4 and 5, the presence of the Ampligase during the SP incubation greatly enhanced the SP activity. In lanes 8−13, various combinations of SP and Taq DNA polymerase, with or without the DNA ligase present, stimulated the SP activity. All incubations were performed in buffer C, and dNTP was absent in these incubations. Comparing lanes 13 and 3, one can see that the SP stimulation by DNA polymerase and DNA ligase is over 20-fold. Exo$^-$ Klenow DNA polymerase

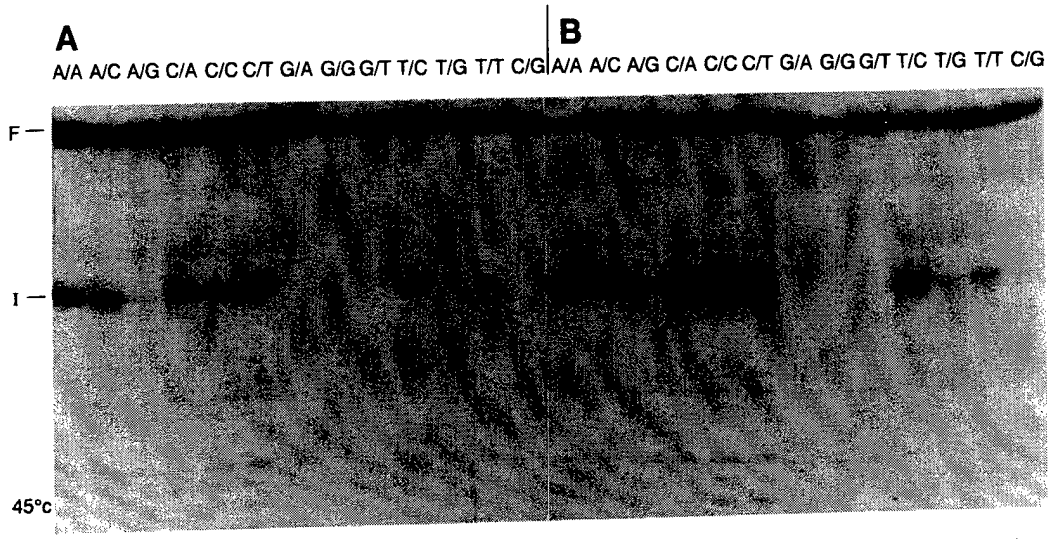


FIGURE 7: Recognition of base-substitution mismatches by SP. 5′-Labeled substrates described in Figure 1C were incubated with 5 ng of SP for 30 min at 45 °C. Samples were analyzed as described in Figure 2. Panel B is the same as panel A, except for the presence of 0.5 unit of Taq DNA polymerase, but no dNTP. F = position of full-length single-strand DNA. I = incision site at the first phosphodiester bond 3′ of a mismatch.

I fragment missing the 3′ to 5′ exonuclease activity can substitute for the Taq DNA polymerase to stimulate SP activity (data not shown), although a thermostable DNA polymerase is more appropriate at 45 °C. The DNA polymerase and DNA ligase, by themselves or together, do not lead to mismatch nicking (lanes 14−17). This lack of incision by a DNA polymerase on mismatch substrates is in contrast to a Y-type junction that can be nicked by eubacterial DNA polymerases (*20*).

The incision of SP at various base-substitutions, in the absence or presence of stimulation by Taq DNA polymerase, is shown in Figure 7. Some base-substitutions are better substrates for SP than others. To the best of our knowledge, no single-strand-specific nuclease other than the CEL I family (*16*) has been able to make such dramatic mismatch-specific incisions. Guanine residues in base-substitutions (G/A, G/G, G/T, A/G, and T/G) and T/T in our model substrate sequence are not incised by SP appreciably in the absence or presence of stimulation by Taq DNA polymerase.

An interesting possibility, but not the only one, is that the spinach SP nuclease may be a natural variant among the CEL I family of mismatch endonucleases. If this were true, sequence comparison will facilitate the identification of the active site and the elucidation of the parameters that control nucleotide specificity. Furthermore, the absence of the guanine cutting ability in SP is coincident with the presence of cutting at A-T-rich sequences. The latter property is not observed for the CEL I family of nucleases, but is a feature of the S1 family of nucleases and the mung bean nuclease (*21*). Therefore, the properties of SP seem to be intermediate between those of S1-type nucleases and CEL I-type nucleases. The availability of the sequences of these nucleases in the future may shed light on their evolutionary relationships and should clarify why SP cannot cut at most guanine nucleotides.

While the mismatch-removal function of SP, coupled with the proofreading and nick-translation ability of a DNA polymerase, forms an efficient mismatch-removal system in vitro, it is unclear whether mismatch repair is a role for SP in vivo. For example, SP is unable to determine which strand should be preserved as template in the mismatch correction process in vitro. However, its activity is consistent with the characteristics of gene conversion where different species, and different gene regions are known to exhibit unequal amounts of sequence conversion. In such a role, the inability to incise a mismatch at a guanine residue may lead to less gene conversion at some sites.

It was previously shown that SP can incise at pyrimidine $TC_{(6-4)}$ dimers and cisplatin adducts (*12*), suggesting that SP may have a role in DNA repair of these lesions. Those studies were done without using a DNA polymerase or a DNA ligase to stimulate SP. It will be interesting to determine whether the SP incision at these adducts will be more efficient under the conditions established in this paper for mismatch incision.

ACKNOWLEDGMENT

We authors are grateful to Dr. Paul Doetsch of Emory University for a gift of purified SP nuclease and for encouragement to pursue this project. We authors thank Drs. Alfonso Bellacosa and Yoshihiro Matsumoto for critical reading of the manuscript.

REFERENCES

1. Linn, S. M., Lloyd, R. S., and Roberts, R. J., Eds. (1993) in *Nucleases*, Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.
2. Shank, T. E., Rhodes, C., Rigby, P. W. J., and Berg, P. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 989−993.
3. Maekawa, K., Tsunasawa, S., Dibo, G., and Sakiyama, F. (1991) *Eur. J. Biochem. 200*, 651−661.
4. Lacks, S. A. (1981) *J. Biol. Chem. 256*, 2644−2648.
5. Ljungquist, S., and Lindahl, T. (1977) *Nucleic Acids Res. 4*, 2871−2879.
6. Taylor, A. F., and Smith, G. R. (1995) *J. Biol. Chem. 270*, 24459−24467.
7. Anderson, D. G., and Kowalczykowski, S. C. (1997) *Genes Dev. 11*, 571−581.
8. Chow, T. Y.-K., and Resnick, M. A. (1988) *Mol. Gen. Genet. 211*, 41−48.

9. Chow, T. Y.-K., and Fraser, M. (1983) *J. Biol. Chem. 258*, 12010−12018.
10. Ramotar, D., Auchincloss, A. H., and Fraser, M. J. (1987) *J. Biol. Chem. 262*, 425−431.
11. Strickland, J. A., Marzilli, L. G., Puckett, J. M., Jr., and Doetsch, P. W. (1991) *Biochemistry 30*, 9749−9756.
12. Doetsch, P. W., McCray, W. H., Lee, K., Bettler, D. R., and Valenzuela, M. R. L. (1988) *Nucleic Acids Res. 16*, 6935−6952.
13. Kowalski, D., Kroeker, W. D., and Laskowski, M., Sr. (1976) *Biochemistry 15*, 4457−4462.
14. Kroeker, W. D., Kowalski, D., and Laskowski, M., Sr. (1976) *Biochemistry 15*, 4463−4467.
15. Sheflin, L. G., and Kowalski, D. (1984) *Nucleic Acids Res. 12*, 7087−7104.
16. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) *Nucleic Acids Res. 26*, 4597−4602.
17. Loeb, L. A., and Silber, J. R. (1981) *Biochim. Biophys. Acta 656*, 256−264.
18. Kowalski, D., and Sanford, J. P. (1982) *J. Biol. Chem. 257*, 7820−7825.
19. Yeung, A. T., Dinehart, W. J., and Jones, B. K. (1988) *Nucleic Acids Res. 16*, 4539−4554.
20. Lyamichev, V., Brow, M. A. D., and Dahlberg, J. E. (1993) *Science 260*, 778−783.
21. Sung, S., and Laskowski, M., Sr. (1962) *J. Biol. Chem. 237*, 506−511.

BI982318Y

Purification and Characterization of the CEL I Nuclease that has High Specificity for Mismatch*

Bing Yang, Xiao Wen[¶], Catherine A. Oleykowski, C. Glenn Miller, Joanne A. Kulinski, David Besack, Jason A. Yeung[†] and Anthony T. Yeung**

Running Title: CEL I Nuclease

From the Fox Chase Cancer Center, Philadelphia, PA 19111

** To whom correspondence should be sent: Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111. Tel: (215) 728-2488. FAX: (215) 728-3647. Email: AT_Yeung@fccc.edu

[¶] Current address: Du Pont Pharmaceuticals Company, Wilmington, DE 19880

[†] Current address: University of Pennsylvania, Philadelphia PA 19104

SUMMARY

CEL I, isolated from celery, is the first eukaryotic nuclease known that cleaves DNA with high specificity at sites of base-substitution mismatch and DNA distortion. The enzyme requires Mg^{++} and Zn^{++} for activity, with pH optimum at neutral pH. This paper reports its purification, over 33,000 fold, to apparent homogeneity. To correlate protein with activity, the band for the homogeneous CEL I, with and without the removal of its carbohydrate moieties, was extracted from SDS-PAGE, renatured, and shown to have mismatch cutting specificity. Partial amino acid sequence was obtained for about 28% of the CEL I polypeptide, which shows moderate similarity to S1 and P1 nucleases. Yet CEL I differs from these nucleases in substrate specificity. Potential orthologs with higher homology to CEL I were identified, including nucleases putatively encoded by the genes *BFN1* of *Arabidopsis, ZEN1* of *Zinnia*, and *DSA6* of daylily coding for a senescence-inducible protein. We propose that CEL I exemplifies a new family of neutral pH optimum, magnesium-stimulated, mismatch-recognizing nucleases, within the S1 superfamily.

INTRODUCTION

Nucleases are important to many aspects of cellular functions in all organisms (1). Some are highly specialized for DNA recombination, replication and repair while other nucleases are for general nucleic acid degradation. The latter group includes mung bean nuclease (2-4), S1 (5), P1 (6), and the pancreatic DNase I (7). However, the biological functions of many nucleases have yet to be revealed.

We previously reported the discovery of a novel family of neutral pH optimum, Mg^{++}-stimulated, mismatch-specific endonucleases in plants (8). These nucleases are abundant and present in various tissues, including roots, stems, leaves, flowers, and fruits. Zn^{++} was found to be necessary for activity, and the nucleases appear to be mannosyl-glycoproteins by their ability to bind to Concanavalin A (ConA) affinity resin. Mg^{++} is required for efficient DNA nicking at the 3' side of the mismatch nucleotide. One such nuclease from celery, CEL I, was used to develop a fluorescence-based mutation detection assay that is highly effective for insertion/deletion and base-substitution mismatches (8). In this report, we describe the purification of CEL I to homogeneity by a novel procedure. Amino acid sequence of 28% of the enzyme was determined and led to the identification of the homologs of CEL I, and suggesting that CEL I represents a new family of nucleases within the S1 superfamily of structurally related nucleases. Enzymatic comparison of CEL I with the mung bean nuclease is reported in the accompanying paper (9). CEL I is the first nuclease of this new family for which a significant amino acid sequence-enzyme activity relationship has been established.

EXPERIMENTAL PROCEDURES

Materials — Plasmid DNA pUC19 was isolated with the QIAGEN Maxi Kit from DH5α host cells, following the manufacturer's instructions. Calf thymus DNA was obtained from Sigma and purified by repeated cycles of proteinase K digestion and phenol extraction (10). Chromatography resins and columns were purchased from Pharmacia Biotech. Toluidine Blue O and Ponceau S were from Sigma. Endo H_f was from New England Biolabs. Phosphocellulose P11 was from Whatman.

Purification of CEL I — All steps were performed at 4 °C. The nuclease activity was monitored by using a RF-I (Replicative Form I) nicking assay (11).

Step 1: Preparation of the crude extract — 105 kilograms of chilled celery stalks were homogenized with a juice extractor. The juice was collected (total 79.34 L) and adjusted to the composition of Buffer A (100 mM Tris-HCl, pH 7.7, 100 μM PMSF). Solid $(NH_4)_2SO_4$ was slowly added to the juice and gently stirred, to a final concentration of 25% saturation. After 30 min, the suspension was centrifuged at 27,000 x g for 1.5 hours. The supernatant (total 70.56 L) was pooled and the concentration of $(NH_4)_2SO_4$ was adjusted to 80% saturation. After 30 min of stirring, the mixture was centrifuged at 27,000 x g for 2 hours. The pellets were resuspended in Buffer B (0.1 M Tris-HCl, pH 7.7, 0.5 M KCl, 100 μM PMSF) and thoroughly dialyzed against Buffer B.

Step 2: Concanavalin A-Sepharose 4B affinity chromatography — 100 ml of ConA resin (cross-linked with dimethylsuberimidate) was added to the 7.71 L sample in bottles that were gently rolled overnight. The resin was packed into a 2.5 cm diameter column. The flow-through fraction, containing no CEL I activity, was discarded. CEL I was eluted at 4 °C by 200 ml of Buffer B containing 0.3 M α-methyl-mannoside. The elution step was repeated 10 more times until no more nuclease activity could be eluted. The elutate was combined and dialyzed against buffer C (50 mM Tris-HCl, pH 8.0, 5 mM α-methyl-mannoside, 0.01% Triton X-100, and 100 μ M PMSF).

Step 3: DEAE-Sephacel chromatography — The dialyzed sample from step 2 (total 2.5 L) was applied to a 400 ml DEAE-Sephacel column of 5 cm diameter previously equilibrated with Buffer C. The subsequent steps were performed using FPLC. The column was washed with 400 ml of Buffer C. CEL I was eluted with a 1 L linear gradient of 10 mM to 1 M KCl in buffer C containing 50 mM α-methyl-mannoside at a flow rate of 5 ml/min, followed by 400 ml of Buffer C containing 1 M KCl and 50 mM α-methyl-mannoside at a flow rate of 8 ml/min. The most active CEL I fractions were pooled and dialyzed against Buffer D (25 mM potassium phosphate, pH 7.0, 5 mM α-methyl-mannoside, 0.01 % Triton X-100, and 100 μM PMSF).

Step 4: Phosphocellulose P-11 chromatography — The dialyzed CEL I pool from step 3 (120 ml) was applied to a 5 cm diameter column packed with 400 ml of P-11 resin. The column was previously equilibrated with Buffer D at a flow rate of 5 ml/min. After sample loading, the column was washed with 625 ml of buffer D containing 50 mM α-methyl-mannoside at a flow rate of 5 ml/min. CEL I was eluted with a 800 ml linear gradient of 20 mM KCl to 1 M KCl in

Buffer D containing 50 mM α-methyl-mannoside at a flow rate of 5 ml/min. The column was further washed with 400 ml of Buffer D containing 1 M KCl and 50 mM α-methyl-mannoside at a flow rate of 8 ml/min. The most active fractions were pooled and dialyzed against Buffer E (50 mM potassium phosphate, pH 7.0, 5 mM α-methyl-mannoside, 0.01 % Triton X-100, and 100 μ M PMSF) containing 1.5 M $(NH_4)_2SO_4$.

Step 5: Phenyl Sepharose CL-4B chromatography — The dialyzed CEL I pool from step 4 (480 ml) was applied to a 5 cm diameter column packed with 400 ml of Phenyl Sepharose CL-4B. The column was previously equilibrated with Buffer E containing 1.5 M $(NH_4)_2SO_4$ at a flow rate of 5 ml/min. After sample application, the column was washed with 400 ml of Buffer E containing 1.5 M $(NH_4)_2SO_4$ and 50 mM α-methyl-mannoside at a flow rate of 5 ml/min. CEL I was eluted from the column with a 500 ml linear reversed salt gradient from 1.5 M to 0 M $(NH_4)_2SO_4$ in Buffer E containing 50 mM α-methyl-mannoside at a flow rate of 5 ml/min. The most active fractions were pooled and dialyzed against Buffer F (50 mM Tris-HCl, pH 8.0, 5 mM α-methyl-mannoside, 0.01 % Triton X-100, and 100 μM PMSF).

Step 6: Mono Q anion-exchange chromatography — A Pharmacia prepacked Mono Q HR 16/10 column was thoroughly washed and equilibrated with Buffer F. The dialyzed CEL I pool from step 5 (336 ml) was applied at a flow rate of 5 ml/min followed by 100 ml of Buffer F containing 50 mM α-methyl-mannoside at a flow rate of 10 ml/min. CEL I was eluted with a 250 ml linear gradient of 0 – 1 M KCl in Buffer F containing 50 mM α-methyl-mannoside at 2 ml/min.

Step 7: Superdex 75 size-exclusion chromatography using the SMART system — The active fractions of step 6, fraction 11 and 12, were combined and concentrated by using Centricon 3 centrifugal concentrators. Aliquots of the concentrated enzyme were applied to a prepacked Superdex 75 PC 3.2/30 column equilibrated with Buffer G (50 mM Tris-HCl, pH 8.0, 100 mM KCl, 10 μM $ZnCl_2$, 0.01 % Triton X-100, and 100 μM PMSF) containing 50 mM α-methyl-mannoside. Five ml of Buffer G containing 50 mM α-methyl-mannoside was used to elute CEL I at a flow rate of 0.05 ml/min. The purity of the active fractions was checked by SDS-PAGE. When additional protein bands were present, the fractions were pooled, concentrated, and purified again using the same size exclusion chromatography until CEL I reached apparent homogeneity.

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) — Polyacrylamide gel electrophoresis in SDS was carried out as previously described (8). Protein bands were detected by using the Gelcode Blue Stain Reagent (Pierce). Molecular weights of the protein bands were determined by using the semi-logarithmic plot of the molecular weights of protein standards versus their relative electrophoretic mobilities.

Endo H_f Removal of N-linked oligosaccharides from CEL I — CEL I sample was denatured in 0.5% SDS at 100 °C for 10 min. Appropriate amount of Endo H_f was added and the reaction was incubated in G5 buffer (50 mM Sodium Citrate, pH 5.5) at 37 °C overnight.

Plasmid DNA RF-I Nicking Assay — The 19.5 μl reaction mixture was in 1X G buffer (20 mM Tris-HCl, 25 mM KCl, and 10 mM $MgCl_2$, pH 7.5) containing 0.5 μg of pUC19 plasmid DNA

and 0.1% Triton X-100. After 0.5 µl of CEL I sample was added, the reaction mixture was incubated at 37 °C for 30 min. The reaction was stopped by adding 10 µl of a 3X stop solution (50 mM Tris-HCl, pH 6.8, 3% SDS, 4.5% β-mercaptoethanol, 30% glycerol, and 0.001% Bromophenol Blue). Ten µl of the mixture was loaded into a sample well of a 0.8 percent agarose gel. After electrophoretic resolution, the DNA was visualized by ethidium bromide staining.

Activity Gel Analysis — This method is a modification of a previously described procedure (12-13). Separating gel is composed of 375 mM Tris-HCl, pH 8.8, 12% acrylamide/0.32% bisacrylamide, 0.1% SDS, 1% glycerol, and 0.75 mg of calf thymus DNA previously denatured by boiling for 10 min. Stacking gel is composed of 125 mM Tris-HCl, pH 6.8, 3.9% acrylamide/ 0.104% bisacrylamide, and 0.1% SDS. Electrophoresis and subsequent steps were performed at room temperature. After the electrophoresis was completed, the gel was washed three times with 10 mM Tris-HCl, pH 7.4, 25% isopropanol, for 20 min each. The gel was then washed with 10 mM Tris-HCl, pH 7.4, for 10 min, and further incubated in 40 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 5mM $CaCl_2$ and 2 µM $ZnCl_2$ overnight. The gel was next incubated in the same buffer at 37 °C for 1 hour, washed in 10 mM Tris-HCl, pH 7.4 for 10 min, and stained with 0.1% Toluidine Blue O in 10 mM Tris-HCl, pH 7.4 for 10 min. Destaining was in 10 mM Tris-HCl, pH 7.4.

Renaturation of CEL I from SDS-PAGE — This method is a modification of a procedure previously described (14). The CEL I fractions were loaded onto the SDS-PAGE (15) in two consecutive lanes. After electrophoresis, the gel was split between the two lanes. One half of the gel was stained with Gelcode Blue Stain Reagent (Pierce) and then aligned with the other half

that was not stained. The gel slice corresponding to the CEL I band in the unstained gel was excised and eluted using an AMICON model 57005 electroeluter, for 2 hours at 20 mA per sample, using the elution buffer (50 mM Tris-HCl, pH 7.5, 180 mM NaCl, 0.1% SDS, 0.1 mg/ml BSA). After elution, the sample was concentrated by using a Centricon 3 unit. Centrifugation was overnight at 7,000 x g. The volume of the sample was measured and 4 volumes of distilled acetone (-20 °C) was added. The sample was incubated in dry ice-ethanol bath for 30 min and then centrifuged at 14,000 x g for 10 min. The precipitated proteins were washed with a buffer consisting of 20% Dilution and Renaturation Solution (50 mM Tris-HCl, pH 7.5, 10% Glycerol, 100 mM NaCl, 10 ml $MgCl_2$, 5 mM $CaCl_2$, 2 μM $ZnCl_2$ and 0.1 mg/ml BSA) and 80% acetone. The sample was precipitated again at 14,000 x g for 10 min. The supernatant was discarded. The residual acetone was decanted by inverting the tube for 10 min. The pellet was air dried for at least 10 min. Twenty μl of Renaturation Solution (6 M Guanidine-HCl, 50 mM Tris-HCl pH 7.5, 10% Glycerol, 100 mM NaCl, 10 ml $MgCl_2$, 5 mM $CaCl_2$, 2 μM $ZnCl_2$ and 0.1 mg/ml BSA) was then used to dissolve the pellet. After 20 min of incubation at room temperature, 1 ml of Dilution and Renaturation Solution was added and the protein was further renatured at room temperature for 12 hours.

Mismatch endonuclease assay — The mismatch endonuclease assay was performed as previously described (8). Briefly, PCR products were amplified using genomic DNA from two individuals, one being wild-type and the other being heterozygous for C insertion in exon 20 in the *BRCA1* gene. The forward primer was 5'-labeled with 6-FAM (blue) and the reverse primer was 5'-labeled with TET (green). The location of the insert in the *BRCA1* gene is 5382 nt position. The resulting heteroduplexes provide 402 bp PCR products containing an extrahelical C or an

extrahelical G. 50 ng of the fluorescently labeled substrate was incubated with CEL I for 30 min at 45 °C in a reaction volume of 20 µl in 20 mM HEPES pH 7.5, 10 mM KCl, 3 mM $MgCl_2$. The reactions were processed as described (8), loaded onto a denaturing 34 cm well-to-read 6 % polyacrylamide gel on an ABI 377 DNA Sequencer and analyzed using GeneScan 3.1 software (Perkin-Elmer). The results are displayed as a gel image.

Preparation of the CEL I Sample for Sequencing — The purified CEL I sample was subjected to 10% SDS-PAGE analysis. After electrophoresis, the protein in the gel was electrophoretically transferred to an Immobilon-PSQ PVDF membrane by using a Western transfer apparatus (Novex). The transfer buffer contained 12 mM Trizma base, 96 mM glycine, and 20% methanol. The transfer condition was 1 hour at 25V (constant voltage). The membrane was next washed extensively with water, and stained with Ponceau S. The CEL I band was excised, destained with water, and sent to the Protein/DNA Technology Center of Rockefeller University for N-terminal and internal peptide micro-sequencing by automated Edman degradation reaction. The N-terminal sequence was determined first (16). The remaining protein fractions were digested with either Trypsin or GluC. The digested peptides were purified by HPLC, and sequenced with Edman Degradation (17).

RESULTS

Purification of CEL I — CEL I was successfully purified to homogeneity, more than 33,000 fold over its specific activity in the buffered celery juice. Table 1 summarizes the purification of CEL I from 105 Kg of celery stalks. To ascertain that the novel activity of CEL I is not the result of a proteolytic degradation of a larger protein, activity gel assays were conducted to analyze the overall nuclease bands at each purification step (Fig 1).

After the 25% $(NH_4)_2SO_4$ precipitation, nucleases less than 39 KDa, as judged by activity gel analysis, were removed (Fig. 1). CEL I eluted from ConA-Sepharose 4B was contaminated with other glycoproteins that bind to ConA. In our previous attempts to purify CEL I (8) and the *Arabidopsis* ortholog ARA I (data not shown), many glycoproteins bands copurified in several chromatography steps, suggesting a possibility of protein aggregation. In the current CEL I purification protocol, two reagents are used to prevent protein aggregation. First, 0.01% Triton X-100 is used in all steps after the ConA column elution. Secondly, α-methyl-mannoside is also included in all buffers after the ConA step to disrupt interactions between CEL I and endogenous lectins that may mediate the aggregation process. Under these new buffer conditions, CEL I was found to resolve from other contaminants in the subsequent chromatography steps using DEAE-Sephacel, phosphocellulose P-11, Phenol Sepharose CL-4B, and Mono-Q, respectively. There are two nuclease bands that copurify during all the purification steps. We show below that the minor band is not derived from the major band. The major nuclease activity, designated CEL I, migrates at 43 Kda on SDS PAGE. The minor activity at 39 KDa is a putative isozyme we named CEL II, also capable of cutting at mismatches.

While the Mono-Q fractions 9 through 15 contained the most active CEL I protein as judged by the plasmid nicking assay, the fractions 11 and 12 were found to contain the least contaminants as judged by the SDS-PAGE (data not shown). These two fractions were combined and CEL I was further purified by using the Superdex-75 PC 3.2/30 (Pharmacia) size-exclusion chromatography with the SMART system (Pharmacia). The purity of the active fractions was examined using SDS-PAGE. The most active fractions were further purified by repeating the size exclusion chromatography procedure. Two final CEL I fractions were obtained. One fraction contained only the major 43 KDa protein designated as CEL I (Fig. 2A). The other fraction contained both the 43 KDa protein and the 39 KDa protein which we designated as CEL II (Fig. 2B, lane 3).

After minimizing the N-linked oligosaccharides by Endo H_f, the 43 KDa major celery nuclease band shifted to the 29 KDa position (Fig. 2B & C, lanes 4) and the 39 KDa minor celery nuclease band shifted to the 37 KDa position (Fig. 2C, lane 4). If CEL II were a degradation product of CEL I, after endo H_f treatment, its polypeptide length should be equal or less than 29 KDa. Amino acid sequence data on CEL II in the future will confirm whether CEL I and CEL II are two different enzymes.

Effects of Reducing Agents on CEL I —It is known that reducing agents can reveal a nick in the enzyme backbone of the Mung Bean nuclease in SDS gel analysis by resolving about 70% of the full length polypeptide into two shorter polypeptides (18). When 1 % β-mercaptoethanol was used in the sample buffer for SDS-PAGE analysis of the CEL I band, CEL I was shifted upward

(Fig 2D, lane 2) but intact. DTT was also tested and similar results were obtained (data not shown). The simplest interpretation is that the CEL I polypeptide does not contain any breakage in the backbone. Instead, disulfide bonds were broken that resulted in the enzyme becoming more extended in the reduced state, and hence slower in electrophoretic mobility. This conclusion is consistent with previous finding that β-mercaptoethanol or DTT greatly reduces the activity of CEL I (8).

Renaturation of homogeneous CEL I and CEL II — Individual celery nuclease bands were excised from the 10% SDS-PAGE and eluted as described in Experimental Procedures. These bands included the 43 KDa band, the 39 KDa band, and their corresponding bands after the Endo H_f digestion. The eluted enzyme fractions were concentrated and renatured. Plasmid nicking assays were carried out to show that the renatured samples were all active nucleases (data not shown). The renatured CEL I before or after Endo H_f digestion and CEL II after Endo H_f digestion were able to incise DNA at a mismatch substrate (Fig. 3). In this experiment, the mismatch incised is a G residue insertion. This experiment is necessarily qualitative because of the uncertainties in the recovery of proteins and activity in the gel elution and renaturation steps. However, the data strengthens the conclusion that CEL I and CEL II are homogeneous and each able to incise at a DNA mismatch, and that most of the carbohydrates on CEL I and CEL II are not essential for activity.

Isoelectric point of CEL I and CEL II — A sample of CEL I, containing a small amount of CEL II, was loaded onto an isoelectric focusing gel (pH 3-10, from Novex). After the gel was stained, the pI of the CEL I and CEL II were obtained by comparison with the standards (Bio-Rad). The

pI of the CEL I band was between 6.0 and 6.5, and the pI of the CEL II band was between 6.5 and 6.8 (data not shown).

Amino acid sequence of CEL I — The amino acid sequence of the N-terminal and three other internal proteolytic peptides of CEL I are shown in Fig. 4. The 72 amino acids identified represent about 28% of the CEL I polypeptide. The ClustalW alignment (19) of these peptides with homologs form the Genbank database is shown in Fig. 5.

DISCUSSIONS

The purification of glycoproteins —We previously described a purification protocol that produced highly enriched CEL I, but was unable to provide the enzyme as a single band on a SDS PAGE gel (8). In the present work, we traced the problem to the presence of endogenous lectins in plant tissues (unpublished data). Such lectins become a problem when the glycoprotein to be purified is less abundant than the lectins. The presence of mannose in the present protocol has overcome this obstacle and has provided a homogeneous preparation of CEL I.

The CEL I obtained by this protocol is surprisingly homogeneous, being a single-band after prolonged migration on a SDS gel (Figs. 2A). Prior to this protocol, purified CEL I appeared as multiple very closely spaced bands. Moreover, after reduction of the new homogeneous enzyme by β-mercaptoethanol or dithiothreitol, the protein is still a single-band, indicating the absence of breakage in the polypeptide backbone. Therefore, previous observation of multiple bands may be due to proteolysis or the presence of multiple glycoforms of the CEL I protein. That the new protocol produces a uniform intact CEL I polypeptide indicates that during previous purification, low levels of proteases and/or glycoprotein modification enzymes was present in aggregations with CEL I. These modification enzymes can act on CEL I when the components are brought into close proximity by the endogenous multivalent lectins in the absence of α methyl-mannoside in the buffers. The lack of backbone breakage in CEL I is in contrast to both the mung bean nuclease (18) and the tobacco extracellular nuclease (20), which each contains one breakage in the polypeptide backbone.

The identification of CEL I—Assuming that the 29 KDa CEL I polypeptide is about 263 residues, similar in polypeptide length to the homologs, the 72 amino acids in Fig. 4 provide the sequence of 28% of the CEL I protein. Extensive alignment of the 72 amino acids in Fig. 5 occurs both at the N-terminal and near the C-terminal of the homolog proteins. Knowing that the purified CEL I polypeptide is of a similar size to the polypeptides coded by the homologs, the extensive peptide sequence identity provides confidence in the conclusion that CEL I is an ortholog of DSA6 (21), BFN1 (22) and ZEN1 (23). That is, they are likely to have the same enzymatic activity.

It can be assumed that CEL I residue 41 is a C because it is invariant among the plant enzymes. The extent of identity among any two homologs of the group #3-#6 of Fig. 5, CEL I, daylily senescence protein DSA6 (21), BFN1 (22) and ZEN1 (23), is very high. For example, there is 58/77 = 75% identity between CEL I and ZEN1, making them likely to be orthologs of each other. Similarly, the orthologs S1 (24) and P1 (6) showed 39/85 = 46% identity in this sequence regions. However, there is only 13/77 = 17% identity between CEL I and S1 nuclease, and 18/77 = 23% identity between CEL I and P1 nuclease. Therefore the sequence information suggests that DSA6, BFN1 and ZEN1 are unlikely to be orthologs of S1 and P1. BFN1 protein has not been purified. ZEN1 nuclease has been purified but only a sequence of seven amino acid of the N-terminal was obtained by Edman degradation protein sequencing (25). In fact, that the second amino acid of ZEN1, a serine, was not determined with confidence suggests the possibility that the ZEN1 sample may contain a mixture of both ZEN1 and its homolog nucZe2 which contains a serine residue at this position (Fig 4). However, as a result of the high degree of sequence identity of DSA6, BFN1 and ZEN1 with CEL I, there is a good probability that the enzymology-amino

acid sequence relation we established for CEL I may be extended to include DSA6, BFN1, and ZEN1 as orthologs.

Fig. 5 also shows four other putative plant nucleases (#7-#10) that include two ZEN1 homologs from *Zinnia* (accession numbers U90265, and U90266), BEN1 of barley (23, accession number D83178), and a putative homologs of *BFN1* of *Arabidopsis* deduced from the genomic sequence (accession number AL022603 emb|CAA18724|). These sequences may be related more to each other than to the sequences of the CEL I orthologs. Future protein expression experiments of the recombinant clones will allow us to test whether these sequences may code for orthologs of the S1 and P1 nucleases or the CEL II endonuclease of celery that can also cut mismatch DNA at neutral pH.

In an alignment of the complete amino acid sequences of all the S1 homologs listed in Fig. 5 (data not shown), the universally conserved residues are the N-terminal tryptophan residue, the five histidine residues, and three aspartate residues, located in different regions of the polypeptide. These nine residues are brought together to bind the three Zn^{++} atoms, as revealed by the X-ray crystallography structure of the P1 nuclease (26-27). The conservation of the catalytic active site suggests that these nucleases share the same mechanism for the cleavage of the phosphodiester bonds, necessitating the conservation of the enzyme structure and scaffold to form the catalytic domain. The differences in substrate preference may lie in the mechanism of substrate recognition, separate from catalysis, such that S1 family nucleases are specific for single-stranded nucleic acids whereas CEL I shows high specificity for mismatch heteroduplexes. The sequences that enable the recognition of different substrates may reside in amino acid

sequences that are less conserved. The availability of the complete sequence of the CEL I protein in the near future will allow us to model the CEL I sequence onto the structure of the P1 nuclease, thereby shed light on the domains where these differences may be.

An interesting feature of P1 may explain the pH dependence of S1. It was pointed out that the P1 molecule is unique among protein structures for the presence of two pairs of uncompensated carboxylates buried within the protein structure (27). These carboxylate pairs are Asp^{66}–Glu^{128} and Asp^{146}-Asp^{151}. If these carboxylates are protonated inside the protein, as may occur at acidic pH, they may not disrupt the protein structure. However, their repulsion without compensation may be one reason for the low activity of P1 at neutral pH, perhaps as a means of preventing toxicity to the cell prior to secretion. P1 and S1 nucleases are 47% in amino acid sequence identity and similar in enzyme properties. The S1 nuclease has conserved only the first of the two pairs of uncompensated carboxylates. The CEL I family of nuclease has conserved only one unpaired carboxylate, corresponding to the Asp^{66} of P1, without other carboxylates located nearby. The mung bean nuclease (MBN) is potentially a member of the plant S1 homolog superfamily. Like S1, it has an acidic pH optimum, is a mannosyl-glycoprotein of 39 KDa, requires Zn^{++} for activity, digests single-stranded DNA and RNA, and a 3' nucleotidase. It is not known whether MBN has buried uncompensated carboxylates. If the sequence of MBN is known, it will shed light on the relevance of these uncompensated buried carboxylates on the pH optimum and stability of these enzymes.

The biological role of the CEL I orthologs is presently unknown. Prior to this study, all S1 homologs were thought to have the same type of activity, i.e. mannosyl-glycoproteins of about 39

KDa, digest DNA and RNA, single-strand specific, acidic pH optimum, and possess 3' nucleotidase activity (1). With the discovery that the preferred substrate of CEL I is a DNA heteroduplex containing a mismatch, or a distortion such as in a supercoiled plasmid, both at neutral pH optimum, the CEL I orthologs have the potential to assume a role other than DNA degradation in lysosomes and vacuoles. The mRNA of the homolog *DSA6* in daylily is induced by 111 fold during senescence (21). However, we have demonstrated that CEL I orthologs are constitutively present in all types of plant tissues that are not undergoing senescence, root, stem, leaves, flowers and fruits (8). Therefore, these enzymes may play a constitutive role in normal plant life as well as inducible to high levels as required under specific circumstances. Consistent with the ability to cut at DNA distortions and mismatches is the speculation that CEL I may be used for antiviral defense or for the processing of the double-stranded RNA signals used for gene regulation in plants (28). With the future cloning and expression of CEL I, it will be feasible to design experiments to examine these potential biological functions. These ortholog sequences might be used to make recombinant proteins and allow the use of site-directed mutagenesis to reveal their enzymatic and biological functions. Antibodies to CEL I and its orthologs will also allow us to identify the locations of these enzymes in the plant and in the cell. If the recombinant BFN1 is shown to possess the same properties as CEL I, it may be identical to ARA I, the CEL I ortholog that we have partially purified from *Arabidopsis* (data not shown). Then, *Arabidopsis* genetics may assist in revealing the roles of these nucleases in the cell.

ACKNOWLEDGMENTS

We thank Drs. Bellacosa and Matsumoto for critical reading of this manuscript, and Dr. Joseph Fernedez for lending us his expertise in micro-sequencing. Protein sequence determination was performed by the Protein/DNA Technology Center of the Rockefeller University.

FOOTNOTES

* This work was supported, in part, by grants NIH CA71426 and U.S. Army DMAD17-97-1-7286 to ATY, by institutional grants from the National Institute of Health to Fox Chase Cancer Center (CA06927, RR05539), and an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The amino acid sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) xxxxxxx.

ABBREVIATIONS

The abbreviations used are: ARAI, mismatch nuclease I from *Arabidopsis*.

bp, basepairs; CEL I, mismatch nuclease I from celery; ConA, Concanavalin A lectin; MBN, mung bean nuclease; nt, nucleotide; PCR, polymerase chain reaction.

REFERENCES

1. Nucleases, eds. Linn, S. M., Lloyd, R. S., and Roberts, R. J. Cold Spring Harbor Laboratory Press, 1993.
2. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4457-4462
3. Sung, S., and Laskowski, M., Sr. (1962) *J. Biol. Chem.* **237**, 506-511
4. Kowalski, D., Natale, D.A. and Eddy, M.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9464-9468
5. Shank, T. E., Rhodes, C. Rigby, P. W. J., and Berg, P. (1975) *Proc. Nat. Acad. Sci. USA*, **72**, 989-993
6. Maekawa, K., Tsunasawa, S., Dibo, G., and Sakiyama, F. (1991) *Eur. J. Biochem.* **200**, 651-661
7. Lacks, S.A. (1981) *J. Biol. Chem.* **256**, 2644-2648
8. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) *Nucleic Acids Research,* **26**, 4597-4602
9. Kodali, N.S., Oleykowski, C.A., Kowalski, D., Yang, B., Miller, C.G., Besack, D.A., Kulinski, J.A., and Yeung, A.T. (1999) Submitted for publication.
10. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
11. Yeung, A.T., Mattes, W.B., Oh, E.Y., and Grossman, L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6157-6161

12. Blank, A., Silber, J. R., Thelen, M. P., and Dekker, C. A. (1983) *Anal. Biochem.* **135**, 423-430

13. Hager, D. A., and Burgess, R. R. (1980) *Anal. Biochem.* **109**, 76-86

14. Kennedy, J. F., Robertson, E. R. (1996) *Bioseparation* 6, 1-15

15. Laemmli, U. K. (1970) *Nature* **227**, 680-685

16. Fernandez, J., Gharahdaghi, F., and Mische, S. M. (1998) *Electrophoresis.* ***19***, 1036-1045

17. Fernandez, J., Andrews, L., and Mische, S. M. (1994) *Anal Biochem.* ***218***, 112-117

18. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4457-4462

19. ClustalW Multiple Sequence Alignment. Kim C. Worley, Human Genome Center Baylor College of Medicine. (http://dot.imgen.bcm.tmc.edu:9331/cgi-bin/multi-align/multi-align.pl).

20. Oleson, A.E., Janski, A.M., Fahrlander, P. D., and Wiesner, T. A. *Arch. Biochem. Biophys.* **216**, 223-233

21. Panavas, T., Pikula, A., Reid, P.D., Rubinstein, B., and Walker, E. L. (1999) *Plant Molecular Biology* **40**, 237-248

22. Abler, M.L., Thompson, D., Lebrasseur, N., van Hoof, A. and Green, P.J. (1999) Genbank accession number U90264

23. Aoyagi, S., Sugiyama, M., and Fukuda, H. (1998) *FEBS Letters* **429**, 134-138

24. Iwamatsu, A., Aoyama, H., Dibo, G., Tsunasawa, S. and Sakiyama, F. (1991) *J. Biochem.* **110**, 151-158

25. Thelen, M.P., and Northcote, D.H. (1989) *Planta* **179**, 181-195

26. Volbeda, A., Lahm, A., Sakiyama, F. and Suck, D. *EMBO J.* (1991) **10**, 1607-1618

27. Romier, C., Dominguez, R., Lahm, A., Dahl, O., and Suck, D. (1998) *Proteins: structure, Function, and Genetics* **32**, 414-424

28. Waterhouse, P.M., Graham, M.W., and Wang, M.B. (1998) *Proc. Nat. Acad. Sci. USA* **95**, 15502-15506

FIGURE LEGENDS

Table I. **Purification of CEL I from celery.** The fold-purification is calculated with respect to the activity measured in the 25% ammonium sulfate supernatant fraction because assays of the crude extract are inaccurate. The SMART system Superose 75 size-exclusion chromatography step is not listed because the fractions from the Mono Q step were purified individually. Protein determination was performed using the Bicinchoninic acid protein assay (Pierce). One unit of CEL I is defined as one thousandth of one single-strand nuclease unit. One unit of single-strand nuclease activity is defined as the amount of enzyme (32 ng CEL I) that produces 1 μg of acid-soluble material at pH 5.5 in 1 min at 37 °C in the absence of Mg^{++} when purified sheared single-stranded calf thymus DNA is used as substrate. Homogeneous CEL I from the Superose 75 size-exclusion chromatography has a specific activity of 3.1E+7 CEL I units/mg, at 33,000 fold purification. 270 units of CEL I will nick 1 μg of supercoiled pUC19 plasmid RF-I at pH 7.5 in 30 min at 37 °C. The pUC19 RF-I nicking assay was used to quantify CEL I throughout purification. That this assay reflects the proper location of CEL I was confirmed by mismatch incision assays. In contrast, the mismatch specificity of CEL I is evident in that 0.3-0.6 units (10-20 pg) of CEL I is optimal in a CEL I mutation detection Genescan assay.

Fig. 1. **Polyacrylamide activity gel analysis of the CEL I purification fractions.** Aliquots of CEL I with approximately equal amounts of CEL I activity from each step of enzyme purification was boiled in SDS gel buffer in the absence of reducing agents, and resolved on a SDS polyacrylamide gel as detailed in the Experimental Procedures. The nucleases, after renaturation, digested the denatured DNA embedding in the gel. The undigested DNA was stained with

Toluidine Blue O to provide a negative image of the positions of the nucleases. Lane 1: molecular weight standards in KDa. Lane 2, Buffered celery juice. Lane 3, 25% ammonium sulfate fractionation supernatant. Lane 4, 80% ammonium sulfate fractionation pellet. Lane 5, sample to ConA-Sepharose column. Lane 6, Eluate from ConA column. Lane 7, Eluate from DEAE-Sephacel column. Lane 8, Eluate from Phosphocellulose P-11 column. Lane 9, Eluate from Phenol Sepharose column. Lane 10, Pool of fractions 11 and 12 from Mono Q column.

Fig. 2. **SDS polyacrylamide gel analysis of purified CEL I and CEL II.** (A): Lane 1, molecular weight standards shown in KDa on the side. Lane 2, 1 μg of homogeneous CEL I enzyme. Panels B and C examine the mobility changes in the CEL I and CEL II protein bands due to EndoH_f treatment. Samples in Panel B contain only CEL I. Samples in Panel C contain a mixture of CEL I and CEL II. Panel D shows the mobility change of homogeneous CEL I after sulfhydryl reduction. The gels were stained with Gelcode Blue. (B) Lane 1, Endo H_f. Lane 2: molecular weight standards. Lane 3, homogeneous CEL I, about 30 ng. Lane 4, CEL I digested with Endo H_f. (C) Lane 1, Endo H_f. Lane 2: molecular weight standards. Lane 3, Purified CEL I with a small amount of CEL II. Lane 4, CEL I and CEL II digested with Endo H_f. (D) Purified CEL I was boiled for 2 min in SDS sample buffer in the presence (lane 2) or absence (lane 3) of 1% β-mercaptoethanol. Lane 1: molecular weight standards. H = Endo H_f, I = CEL I, II = CEL II.

Fig. 3. **Incision at mismatch substrate by CEL I and CEL II proteins renatured from SDS gel.** CEL I and CEL II protein bands were excised from a SDS gel and renatured as described in Experimental Procedures. The renatured enzyme was used to digest a 402 bp fluorescently labeled PCR product of exon 20 of the *BRCA1* gene. Lanes 1-6 are homoduplexes made from

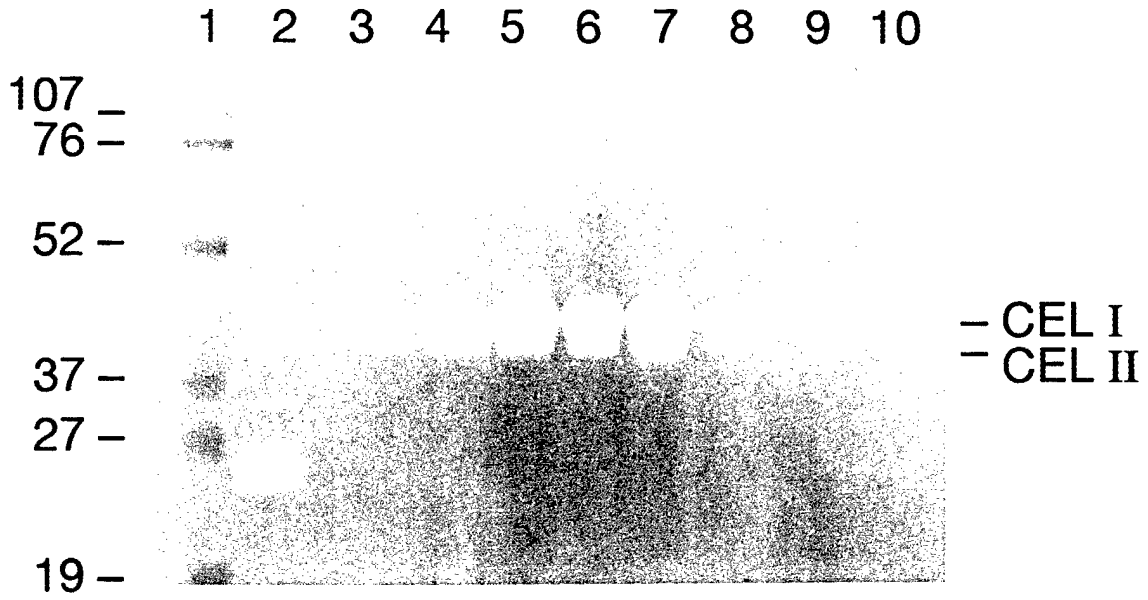
wild type DNA samples containing no mismatch in exon 20. Lanes 7-12, because of the heterozygous nature of this sequence in the sample, the PCR product is a heteroduplex in which one strand contains a G residue insertion. CEL I incision at the 3' side of this extrahelical G residue produces a green band as indicated in the figure. Lanes 1 and 7, substrate with no CEL I treatment. Lanes 2 and 8, incision of the substrate by purified native CEL I. Lanes 3 and 9, incision of substrate by renatured 29 KDa CEL I polypeptide band originated from EndoHf digestion of the 43 KDa CEL I band. Lane 4 and 10, incision of the substrate by the renatured 37 KDa CEL II polypeptide band originated from Endo H_f digestion of the 39 KDa CEL II band. Lanes 5, 6, 11, and 12, incision of the substrate by renatured 43 KDa CEL I band.

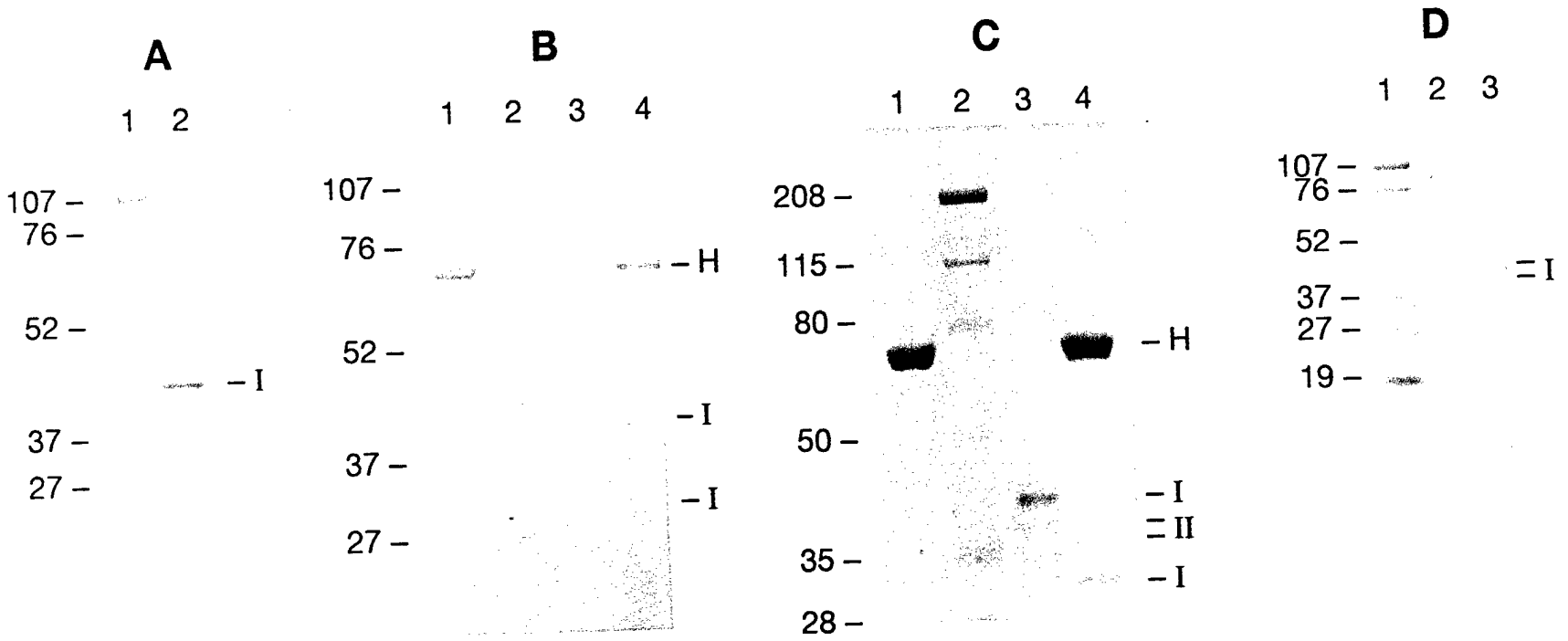
Fig. 4. **Partial amino acid sequence of CEL I.** The amino acid sequence deduced from Edman degradation sequencing of the N-terminal of CEL I and three internal peptides produced by proteolysis are shown. X = unknown residues.

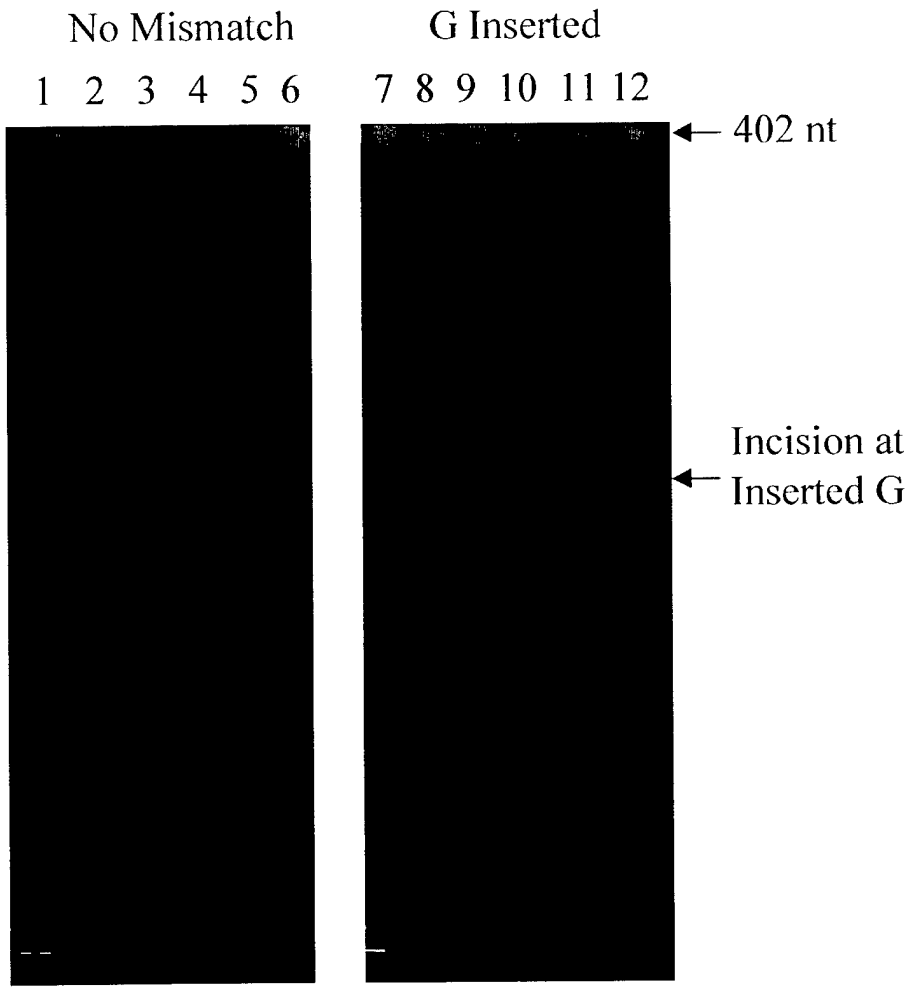
Fig. 5. **ClustalW alignment of the amino acid sequences of CEL I with homologous sequences.** The Genbank accession numbers of the homologous sequences are indicated in brackets: 1: (P24021) Nuclease S1 of *Aspergillus oryzae*; 2: (P24289) Nuclease P1 of *Penicillium citrinum*; 3: The amino acid sequences of CEL I determined by Edman degradation; 4: (AF082031) daylily senescence-associated protein 6 (DSA6) of *Hemerocallis hybrid cultivar*; 5: (U90264) bifunctional nuclease BFN1 of *Arabidopsis thaliana*; 6: (AB003131) ZEN1 endonuclease of *Zinnia elegans*; 7: (U90266) nucZe2 bifunctional nuclease of *Zinnia elegans,*; 8: (U90265) nucZe1 bifunctional nuclease of *Zinnia elegans*; 9: (D83178) BEN1 Barley endonuclease of *Hordeum vulgare*; 10: (AL02263, emb:CAA18723) hypothetical protein of

Arabidopsis thaliana. Identical amino acid residues among homologs numbers 3-6 are highlighted in bold, as are those residues that are similarly conserved in other homologs. The amino acid residue numbers of the P1 nuclease is indicated above the alignments.

Purification Step	Volume in Liter	Total Protein, mg	Total Activity, CEL I units	Specific Activity, units/mg	Protein Fold-purification
Buffered Juice	79.34	19,399	1.89E+7	9.74E+2	
25% $(NH_4)_2SO_4$ supernatant	70.56	17,005	1.57E+7	9.23E+2	1
80% $(NH_4)_2SO_4$ pellet	8	2,072	9.01E+6	4.35E+3	4.5
ConA-Sepharose 4B	2.5	6.75	3.64E+6	5.39E+5	553.8
DEAE-Sephacel	0.12	2.69	2.38E+6	8.83E+5	907.6
Phosphocellulose P-11	0.48	0.408	1.53E+6	3.75E+6	3,854
Phenol Sepharose CL-4B	0.34	0.054	5.61E+5	1.04E+7	10,676
Mono Q	0.03	0.03	3.60E+5	1.20E+7	12,316


1 2 3 4 5 6 7 8 9 10
107
76
52
37
27
19
CEL I
CEL II


A
1 2
107 –
76 –
52 –
– I
37 –
27 –
B
1 2 3 4
107 –
76 –
– H
52 –
– I
37 –
– I
27 –
C
1 2 3 4
208 –
115 –
80 –
– H
50 –
– I
= II
35 –
– I
28 –
D
1 2 3
107 –
76 –
52 –
= I
37 –
27 –
19 –


No Mismatch
G Inserted
1 2 3 4 5 6
7 8 9 10 11 12
402 nt
Incision at
Inserted G

N-terminal sequence:
WSKEGHVMTCQIAQDLLEPEAA
HAVKMLLPDYANGXLSSLXVWP.

Internal peptide from GluC digest:
XSWLQDVE.

Internal peptides from tryptic digest:
CDDISTCANKYAKE.
LACNWGYK.

```
            1                                                   44
 1 S1       WGNLGHETVA Y-IAQSFVASS TESFCQNILG DDSTSYLANV ATWA
 2 P1       WGALGHATVA Y-VAQHYVSPE AASWAQGILG SSSSSYLASI ASWA

 3 CEL I    WSKEGHVMTC Q-IAQDLLEPE AAHAVKMLLP DYANGxLSSL xVWP
 4 DSA6     WSKEGHIVTC R-IAQDLLEPE AAETVRNLLP HYVDGDLSAL CTWP
 5 BFN1     WSKEGHILTC R-IAQNLLEAG PAHVVENLLP DYVKGDLSAL CVWP
 6 ZEN1     WSKEGHVMTC Q-IAQELLSPD AAHAVQMLLP DYVKGNLSAL CVWP

 7 nucZe1   WGVDGHFITC K-IAQGRLSQT AVDAVNSLLP EYAEGDLASL CSWA
 8 nucZe2   WGKEGHYATC K-IAQSFLSEE ALNAVKELLP ETAEGDLASV CSWP
 9 BEN1     WGKEGHYMTC K-IADGFLTSE ALTGVKALLP SWANGELAEV CSWA
10 CAA18723 ISSLGYPLWR RDLRKSYFEED TVVAVKKLLP ESANGELAAV CSWP

            182                                           222
 1 S1       TGTYSSKKD SWTDGIDIKD PVSTSMIWAA DANTYVCSTV LD
 2 P1       SGNYTAQAI GWIKGDNISE PITTATRWAS DANALVCTVV MP

 3 CEL I    ExSWLQDVE xxxxCD---D ISTCANKYAK ExxxLACNWG YK
 4 DSA6     TGIWSDDTS SWGECD---D LFSCPKKWAS ESISLACKWG YK
 5 BFN1     NGLWHDDLS SWTECN---D LIACPHKYAS ESIKLACKWG YK
 6 ZEN1     HGLWSDDVN SWKDCD---D ISNCVNKYAK ESIALACKWG YE

 7 nucZe1   N-VWGDQVK AWENCS--AN QKTCPNIYAT EGIKAACNWA YK
 8 nucZe2   D-RWSNDIS SWVNCT--SG EEVCPDPWAS ESIKYSCNYA YR
 9 BEN1     D-DWSSEEK QWETCR--SK TTTCAEKYAQ ESAVLACD-A YE
10 CAA18723 N-GWSNDVP SWESCQ--LN QTACPNPYAS ESIDLACKYA YR
```

The CEL I Nuclease is Catalytically Different from the Mung Bean Nuclease, Two Nucleases of the S1 Superfamiliy*

Nagendra S. Kodali[§], Catherine A. Oleykowski[§], David Kowalski[¶], Bing Yang, C. Glenn Miller, David Besack, Joanne A. Kulinski, and Anthony T. Yeung**

Running title: CEL I Nuclease versus Mung Bean Nuclease

From the Fox Chase Cancer Center, Philadelphia, PA 19111

[¶] the Roswell Park Cancer Institute, Buffalo, New York 14263.

[§] These authors contributed equally to the work in this report.

** To whom correspondence should be sent: Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111. Tel: (215) 728-2488. FAX: (215) 728-3647. Email: AT_Yeung@fccc.edu

SUMMARY

CEL I, a novel nuclease purified from celery, nicks a DNA heteroduplex at the site of a base-substitution mismatch efficiently at neutral pH. Surprisingly, it has significant sequence homology to S1 and P1 nucleases. Because S1 and P1 are almost identical in properties to the mung bean nuclease, the latter is also likely to be a homolog of CEL I. In this paper, we compared the homogeneous CEL I nuclease with two reliable sources of homogeneous mung bean nuclease to establish that these enzymes, although physically similar, are catalytically distinct. CEL I cuts DNA at a base-substitution while mung bean nuclease cannot. Although mismatch recognition is often attributed to single-strandedness, CEL I is about 32 times less active on denatured DNA than the mung bean nuclease. This suggests that CEL I is not just another single-strand specific nuclease and single-strandedness is not the major feature it recognizes in a mismatch. In the course of these studies, we made the unexpected observation that mung bean nuclease, at neutral pH, is vastly stronger as an RNase than a DNase. We propose that CEL I and mung bean nuclease represent members of two different families within a structurally related S1 superfamily.

INTRODUCTION

S1 (2), P1 (3) and mung bean nuclease (MBN) (4-6) belong to a family of nucleases generally described as the S1 family (1). These nucleases are Zn^{++} metalloenzymes of about 39 KDa, and are mannosyl glycoproteins. They digest both DNA and RNA. These nucleases are highly active at pH 4.5 to 5.5, and mostly inactive at neutral pH range. This property is consistent with their preference, by several thousand fold, for single-stranded nucleic acids versus double-stranded nucleic acids because double-strandedness is less well maintained at acidic pH. The single-strand specificity led to the use of these nucleases for mapping single-stranded regions in nucleic acid duplexes (7), such as in the S1 nuclease protection assay (8), and in the preparation of blunt-ended DNA in some recombinant DNA experiments (9). Efforts to apply them to the detection of base-substitution mutations have generally been unsuccessful (10).

CEL I is a nuclease we discovered in celery. It nicks one strand of a DNA heteroduplex at the 3' side of the mismatch (11). Celery contains as much as 40 μg of psoralen per gram of tissue (12). Such intercalators may be expected to cause frameshift mutations unless they are recognized and removed rapidly. Celery may be expected to have a system for preventing frameshift mutations. CEL I recognizes all base-substitution mismatches as well as extrahelical nucleotides as small as single nucleotide insertions. It has a neutral pH optimum in mismatch incision, and requires both Zn^{++} and Mg^{++} for optimal activity. Although the mismatch specificity of CEL I is suitable for mismatch repair, its biological role(s) is not yet known. However, its ability to detect all the possible mismatches has

enabled an enzymatic mutation detection method that can replace the popular method of single-strand conformation polymorphism (SSCP). CEL I is a mannosyl-glycoprotein of about 39 KDa. It has secondary activity on single-stranded DNA and RNA, raising the possibility that it is related to the S1 family of nucleases, but evolved to be active at neutral pH and to possess different substrate specificity.

We have recently purified CEL I to homogeneity and obtained peptide sequences of 28% of the protein (13). The available CEL I amino acid sequence shows low, but significant, homology (18% identity) to S1 nuclease. In contrast, CEL I is about 76% identical to several putative nuclease genes in plants, namely, the genes that code for the daylily senescence protein DSA6 (14), the *Arabidopsis* hypothetical bifunctional nuclease BFN1 (15), and a Zinnia hypothetical nuclease ZEN1 (16-17). A CEL I ortholog activity is also abundant in mung bean, both in the root and in the shoot (11). Incidentally, mung bean root also contains the S1-like MBN (4-6). Thus biochemical evidence suggests that mung bean has at least two S1-like nucleases with different functions. This possibility is consistent with the presence of at least three homologs of S1, in either *Arabidopsis* (accession numbers U90264, AL022603, for emb|CAA18723| and emb|CAA18724|) or Zinnia (accession numbers AB003131, U90265, and U90266), in Genbank. Whether these plant homologs have differences in function may be revealed by a careful comparison of their enzymatic properties.

Plant nucleases with S1-like activities and properties are often called Nuclease I (1). This nomenclature is no longer adequate because a plant may have more than one S1 homolog.

Even if these homologs are structurally similar to S1, they may have different functions. MBN is very similar to S1 in catalytic properties, and may represent the plant orthologs of S1. CEL I and MBN are available in homogeneous preparations for activity comparison. In this paper, we illustrate that CEL I and MBN are different enzymes because of vast differences in catalytic properties.

Experimental Procedures

CEL I nuclease — CEL I was purified to homogeneity as described in the accompanying paper in this issue (13). Protein concentration was measured using the Bicinchoninic acid protein assay (Pierce). One unit of single-strand DNase activity is defined as the amount of enzyme that produces 1 μg of acid-soluble material at pH 5.5 in 1 min at 37 °C in the absence of Mg^{++} when purified sheared single-stranded calf thymus DNA is used as the substrate. This is the same condition used for our unit definition of the MBN. The conditions for the unit definition of MBN from different sources are similar, but not identical. In spite of the fact that the digestion of single-strand DNA is not the major activity of CEL I, and that it is about 30% more active in the presence of Mg^{++}, the single-strand nuclease assay in the absence of Mg^{++} is the easiest method for standardization and comparison of this family of enzymes. One unit of single-strand nuclease activity of CEL I equals 32 ng of homogeneous CEL I. For the practical purpose of the CEL I mutation detection assay described below, 1 CEL I unit is defined as being equal to 1/1000 of one unit of single-strand nuclease activity. 270 CEL I units will nick 1 μg of supercoiled pUC19 plasmid RF-I at pH 7.5 in 30 min at 37 °C.

Sources of Mung Bean Nuclease — MBN was purchased from Pharmacia Biotech, #27-0912, herein called 'MBN-A', or purified as previously described (18), herein called 'MBN-B'. MBN assay conditions and the measurement of protein concentrations vary in different laboratories and may partially influence the quantitation in this study. MBN-A is FPLC purified, homogeneous, with a specific activity of 1.64×10^6 units/mg in the manufacturer's assay conditions, but 1.42×10^6 units/mg in our assay conditions. The

enzyme exhibits a single-band in SDS PAGE. MBN-B is an older preparation of the original MBN and has a specific activity of 4×10^5 units/mg in the assay conditions used in this report. The enzyme appeared as a single band of about 39 KDa on a non-reducing SDS PAGE (data not shown). One unit of MBN-A single-strand DNase activity equals 0.7 ng of enzyme in our assay.

RF-I nicking assay — 1.1 μg of pPK201/cat (a pUC19 plasmid derivative, data not shown with pUC19 are similar) was incubated with the designated amount of MBN or CEL I for 30 min at 37 °C in a volume of 30 μl of Buffer A (20 mM sodium acetate pH 5.5, 10 mM KCl), or Buffer B (20 mM HEPES pH 7.5, 10 mM KCl) in the presence or absence of 3 mM $MgCl_2$. To stop the reaction, 5 μl of stop solution (50 mM Tris-HCl, pH 6.8, 3 % SDS, 4.5 % β-mercaptoethanol, 30 % glycerol, and 0.001 % Bromophenol Blue) was added. 24 μl of the final mixture was loaded onto a 0.8 % agarose gel. After electrophoresis and staining with ethidium bromide, a photograph of the gel was taken and the negative was scanned using the IS-1000 Digital Imaging System (Alpha Innotech Corporation). The RF-I band was quantified using IS-1000 v2.02 software.

Single-strand DNase assay — The DNA solubilization assay was similar to that previously described (19). Fifty μg of heat-denatured calf thymus DNA (Calbiochem # 2618, purified by repeated pronase treatment, phenol extraction and dialysis) was incubated with 0.7 ng of MBN-A, or 1.9 ng of MBN-B, or 16 ng of CEL I, in 100 μl of Buffer A or Buffer B, with or without 3 mM $MgCl_2$. At the designated times, 100 μl of cold 20 mM $LaCl_3$ in 0.2 N HCl was added to stop the reaction. After centrifugation

(21,000 x g, 40 min), the absorbance at 260 nm of the supernatant was measured using a spectrophotometer to determine the amount of DNA that had become acid-soluble.

Mismatch endonuclease assay — The mismatch endonuclease assay was performed as previously described (11). Briefly, PCR products were amplified using genomic DNA of individuals heterozygous for certain alterations in three different exons in the *BRCA1* gene. The forward primer was 5'-labeled with 6-FAM (blue) and the reverse primer was 5'-labeled with TET (green). The location of the mismatches in the *BRCA1* gene are 300 nt, 4184 nt, 4421 nt, and 5382 nt positions. They correspond to a T→G base substitution in exon 5, a 4 base deletion in exon 11, a C→T polymorphism in exon 13, and a C insertion in exon 20, respectively. The four resulting heteroduplexes provide a 235 bp PCR product containing a T/C or a G/A base-substitution mismatch, a 387 bp PCR product containing a 4 base loop, a 323 bp product containing either a C/A or a T/G base-substitution mismatch, and a 402 bp product containing an extrahelical C or an extrahelical G. 50 ng of the fluorescently labeled heteroduplex was incubated with 7 ng of MBN-A, or 11 ng of MBN-B, or 10 pg of CEL I (0.3 units) for 30 min at 37°C or 45 °C in a reaction volume of 20 µl in Buffer B in the presence or absence of 3 mM $MgCl_2$. The reactions were processed as described (11), loaded onto a denaturing 34 cm well-to-read 6 % polyacrylamide gel on an ABI 377 Sequencer and analyzed using GeneScan 3.1 software (Perkin-Elmer). The results can be displayed as a gel image (Fig. 4) or as a display of the peak profile of each lane of the gel image (Fig. 3).

Single-Strand RNase assay — Fifty μg of purified Torula Yeast RNA (Amicon #7120) was incubated with 0.7 ng of MBN-A, or 16 ng of CEL I, in 100 μl of Buffer A or Buffer B, with 3 mM $MgCl_2$ at 37 °C. At the designated times, 13 μl of cold 3M sodium acetate pH 5.2 and 282 μl of ethanol was added. The mixture was put at –20 °C overnight. After centrifugation to precipitate the RNA (21,000 x g, 45 min), the absorbance at 260 nm of the supernatant was measured using a spectrophotometer to determine the amount of RNA that had become soluble.

RESULTS

The RF-I nicking activity of CEL I and MBN — Supercoiled plasmid replicative form I (RF-I) DNA exhibit local regions of instability in the double-helix that can be attacked by nucleases. Upon the first nick, the superhelical stress is relieved, and the DNA is no longer a substrate for most single-strand nucleases. The RF-I nicking activities of MBN and CEL I at pH 5.5 versus pH 7.5 are shown in Fig. 1. Panel A and B compare the nicking of RF-I by MBN-A at the two pH's in the presence and absence of Mg^{++}. In panel A, under condition of initial kinetics, the inhibition of MBN by 3 mM Mg^{++} is about 90%. About 70% of the RF-I is nicked by 7 pg of MBN-A in 30 min at pH 5.5. In panel B, 7 ng of MBN-A can only nick about 20% of the RF-I in 30 min at pH 7.5. Similar result is obtained for MBN-B in panels C and D. Similar comparison of CEL I RF-I nicking activity is shown in panel E for pH 5.5, and panel F for pH 7.5. The data shows that CEL I is about twice as active in RF-I nicking in the presence of Mg^{++} than in the absence of Mg^{++}. Comparing the 5 pg data points, CEL I is twice more active at pH 7.5 than at pH 5.5.

The single-strand DNase activity of CEL I and MBN — The digestion of denatured purified calf thymus DNA by MBN and CEL I are shown in Fig. 2. For ease of comparison, different amounts of MBN and CEL I were used so that the assays are in a similar range of total activity. The amounts of enzyme used for MBN-A, MBN-B, and CEL I were 0.7 ng, 1.9 ng, and 16 ng, respectively. The lack of activity by MBN at pH 7.5 is obvious in panels A and B. The Mg^{++} inhibition of MBN is also observed for the

activity on single-stranded DNA. In contrast, CEL I is more active in the presence of Mg^{++} than in the absence. Importantly, comparing the initial kinetics in panels A and C for the highest activity condition for each enzyme, MBN-A in the absence of Mg^{++} at pH 5.5 appears to be about 32 times higher in single-strand nuclease specific activity than for CEL I in the presence of Mg^{++} at pH 5.5 (1.42×10^6 μg DNA solubilized/min/mg protein versus 4.46×10^4 μg DNA solubilized/min/mg protein).

The mismatch endonuclease activity of CEL I and MBN — The nicking of DNA duplexes containing mismatches by MBN and CEL I are shown in Fig. 3. The mismatch with a four base loop is nicked by CEL I and both preparations of MBN at pH 7.5 (A, B, C). Note the higher amounts of MBN needed in this reaction. However, even at 1000 times more enzyme than CEL I, MBN is unable to specifically nick at base-substitutions at a single base mismatch (D, E, G, and H). When the same amount of MBN protein is incubated with DNA substrates at pH 5.5 as at pH 7.5 the substrate is almost completely digested (data not shown). When a lesser, more appropriate amount of MBN is incubated with the DNA substrate at pH 5.5, no mismatch-specific nicking is seen (data not shown). CEL I nicks at the base-substitution mismatch (panel F) and at the extrahelical nucleotide (panel I). In panel F, the blue peak at position 183 nt corresponds to the nick at the 3' side of the mismatch on the 6-FAM-labeled strand of the heteroduplex, and the green peak at position 142 nt corresponds to the nick at the 3' side of the mismatch on the TET-labeled strand. Some of the other blue peaks are non-specific cutting by CEL I; it is important to note that if one incubates the reaction for a longer time, or with more CEL I enzyme, most of these non-mismatch specific peaks will be removed while the mismatch-specific

peaks will remain (Fig. 4). The reason is that these background bands are often non-specific heteroduplexes of PCR products in which the two DNA strands do not basepair properly. These duplexes are nicked by CEL I at non-specific positions, and their signal becomes diffused. In panel I, the green peak at 252 nt corresponds to the nick at the 3' side of the extrahelical G on the TET-labeled strand of the PCR product. A blue peak corresponding to the nick at the extrahelical C on the 6-FAM-labeled strand is expected at position 151 nt, but is not seen. CEL I may have nicked the 6-FAM-labeled strand near its 5'-end removing the dye, making it unable to score the blue peak in the assay. Alternatively, the insert C substrate may have been out-competed by the insert G substrate.

Mg^{++} and pH dependence of CEL I — A gel-image of the automated DNA sequencer analysis of the CEL I incision at the mismatch of a T→G base substitution is shown in Fig. 4. Lanes 1-4 are mock reactions without CEL I. The full length 235 nt PCR product is seen on top of the image, and imperfect PCR products are seen as the bands dispersed below. In lane 5, in the presence of CEL I, Mg^{++} and pH 7.5, the blue incision band of 156 nt and the green incision band of 80 nt are observed as indicated. In the absence of Mg^{++} or in pH 5.5 (Lanes 6-8), mismatch-specific incisions are not significant. This experiment also illustrates how the imperfect PCR byproducts seen in lanes 1-4 are eliminated by CEL I in lanes 5-8, especially under the conditions of lane 5.

The RNase activity of CEL I and MBN — A property common to S1 and CEL I is the ability to digest both RNA and DNA, a feature referred to as "sugar non-specific" or

"bifunctional" in literature. We have compared the specific activities of MBN and CEL I on RNA using conditions comparable to their DNase activities. The specific questions addressed here are whether the RNase activity is pH-dependent, and whether the specific activities of the RNase and DNase are similar for each enzyme. Our assay measures the digestion of RNA to soluble nucleotides and short RNA fragments. The specific activity of the RNase activity of MBN-A (Fig. 5A) is comparable to its single-strand DNase activity (Fig. 2A). The specific activity of CEL I is 50 times less than MBN-A on Torula Yeast RNA (Fig. 3A) at pH 5.5. This value is consistent with our finding that CEL I is about 32 times lower in specific activity than MBN-A using denatured calf-thymus DNA as substrate. CEL I as an RNase is slightly more active at pH 7.5 than at pH 5.5. This is opposite to the observation for the single-strand DNase activity of CEL I, but the differences are small. Thus MBN at pH 5.5, and CEL I at pH 5.5 and pH 7.5, showed no preference for RNA versus DNA. MBN-A digested RNA at pH 7.5 with the same specific activity as at pH 5.5 (Fig. 5). This is in striking contrast to MBN-A's little to no ability to digest single-stranded DNA at pH 7.5 (Fig. 2A). Similar results were found for the RNase activity of MBN-B (data not shown).

DISCUSSION

The pH dependence of CEL I and Mung Bean Nuclease — In the RF-I of plasmid pUC19, supercoiling induces regions of single-strandedness that can become a substrate for nucleases. Moreover, regions such as the origin of replication are known to form stem-loop structures. It has also been shown that there are destabilized sequences in supercoiled plasmids (20). Although our assay measures the first nicking event in the pUC19 RF-I, it is unclear that the nicks for each enzyme refer to the same sites on the plasmid. Neither is it clear that the nicks occur at the same locations on the plasmid under each pH and Mg^{++} condition. Nonetheless, the RF-I nicking assay is a convenient method to contrast the differences between the MBN and CEL I nucleases. Our data clearly demonstrate that MBN nicks RF-I more quickly at pH 5.5 than at pH 7.5 by more than 1000 fold, yet CEL I is more active at pH 7.5 than at pH 5.5.

The >1000 fold higher activity of MBN at acidic pH on RF-I cutting may be a function of the catalytic mechanism of the enzyme. Another factor that contributes to faster rate of RF-I nicking at acidic pH may be the partial unwinding of a plasmid at acidic pH, thereby producing a greater propensity for single-strandedness. In the case of CEL I being active on plasmid RF-I at neutral pH, one may speculate that a partial unwinding of the RF-I occurs upon the binding of CEL I. Alternatively, CEL I may not be recognizing single-strandedness in the plasmid. The reason is that in spite of CEL I being more active in the digestion of single-stranded DNA at pH 5.5 than at pH 7.5 (Fig. 2), CEL I is less active in RF-I nicking at pH 5.5 than at pH 7.5 (Fig. 1). The possibility that CEL I is recognizing

the structural junction between a single-stranded region and a double-stranded region will be tested in the near future.

When CEL I uses denatured DNA as a substrate, the specific activity of CEL I is 20 times less than MBN-A (Fig. 2C) at acidic pH and only slightly improved at pH 7.5 in the presence of Mg^{++}. In RF-I nicking, which reflects the recognition of destabilized helices, CEL I specific activity is only 2 times less than MBN-A at pH 5.5, but CEL I is 1000 times more active at pH 7.5 (Fig. 1). Moreover, CEL I nicks a mismatch heteroduplex containing four extrahelical bases at 700 times higher specific activity than MBN-A (Fig. 3A, B, C). Lastly, only CEL I can nick DNA at base-substitutions. Therefore, it is evident that CEL I is not primarily a single-strand DNase. Moreover, single-strandedness per se is not what CEL I recognizes in a mismatch substrate.

The role of Mg^{++} in the activity CEL I and the MBN — The initial rate of RF-I nicking by MBN at pH 5.5 is inhibited by Mg^{++} by about 10 to 20 fold. In contrast, CEL I is stimulated by Mg^{++} under all assay conditions. The CEL I nicking of RF-I significantly increases in the presence of Mg^{++} at both pH's. By the RF-I nicking assay itself, it is not possible to distinguish whether the effect of the Mg^{++} is on the plasmid DNA structure or on the enzyme. With single-stranded DNA as substrate, the effect of Mg^{++} on the enzymes was lower perhaps because the effects of Mg^{++} on substrate superhelicity is not involved. With the mutation detection assay, it is clear that Mg^{++} is required for optimal CEL I incision at mismatches in double-stranded DNA (Fig. 4). If CEL I and MBN should use the same catalytic mechanism for phosphodiester bond cleavage, their

differences may lie in how the substrates are recognized. The role of Mg^{++} may lie in substrate recognition and not in DNA hydrolysis.

It is known that Zn^{++} can satisfy both a catalytic role and a structural role in a metalloprotein, while Mg^{++} mainly serves in a structural role (21). It is possible that Mg^{++} serves in CEL I to produce positively charged surfaces at clusters of negatively charged residues so as to facilitate the binding of the double-stranded DNA and the melting of mismatch DNA heteroduplexes. Future experiment after the sequence of CEL I has been revealed will allow this hypothesis to be tested.

Relative activity on mismatch heteroduplexes — Fig. 3 shows the relative activity of CEL I and MBN on three kinds of heteroduplexes. The three enzymes specifically cut at the extrahelical DNA loop of four nucleotides at pH 7.5. The ability of CEL I to cut specifically at base-substitution mismatch and single-nucleotide insertion distinguishes it from MBN. The inability of MBN to cut at base-substitution mismatches is a property similar to that of the S1 nuclease. S1 is known to be able to nick at extrahelical DNA loops of 4 nucleotides or greater, but not at base-substitutions mismatches (10). While S1 and MBN are known to nick at double-stranded DNA at A-T rich regions in the absence of a mismatch, this does not occur for CEL I unless the A-T rich sequence is further destabilized by being at the termini of a DNA duplex (data not shown).

RNase activity — We observed that MBN is primarily an RNase at neutral pH with the RNase activity at least one thousand times greater than the DNase activity. This presents

a challenging system to elucidate both mechanistically and biologically. The pH dependence may lie in the binding step or the catalytic cleavage step. Whether S1 and P1 are active as RNase at neutral pH should also be tested.

Effect of CEL I properties on MBN purification — Given the high activity of the MBN in crude extracts of mung bean, it was surprising to us that the CEL I-like activity could have been detected in the crude mung bean extract (11). The findings in this report provides an explanation for the serendipity in the discovery of CEL I. CEL I was found using the RF-I nicking assay at neutral pH in the presence of Mg^{++}. Under that condition, MBN is inhibited hundreds to thousands of fold by the pH and the Mg^{++}. This combination of properties created a CEL I assay that was essentially devoid of the influence of the MBN and thus enabled us to purify CEL I. The converse is not true for MBN purification. CEL I is active in all the assay conditions used for the MBN, albeit about 10 times less active at pH 5.5 in the single-strand DNA solubilization assay. Thus commercially available MBN preparations that are not homogeneous in SDS PAGE analysis usually contain a variable amount of contamination of a CEL I ortholog detectable with the mismatch detection assay (data not shown).

In summary, we have shown that CEL I and MBN are enzymatically distinct. They appear to represent two families that may have preserved some of the P1 protein structure as a scaffold to support the catalytic center (22), but evolved to have different DNA binding domains on the exterior of this scaffold. Our hypothesis is that the S1 family of nucleases, represented in plants by the MBN, has evolved to accommodate mainly single-stranded

DNA and RNA, while the CEL I family of orthologs is able to accommodate mainly double-stranded DNA at neutral pH, with further evolution to make the base-substitution mismatch a favorable target.

ACKNOWLEDGEMENTS

We thank Drs. Matsumoto and Bellacosa for critical reading of this manuscript.

FOOTNOTES

* This work was supported, in part, by grants NIH CA71426 and U.S. Army DMAD17-97-1-7286 to ATY, GM30614 to DK, by institutional grants from the National Institute of Health to Fox Chase Cancer Center (CA06927, RR05539), and an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

ABBREVIATIONS

The abbreviations used are: ARAI, mismatch nuclease I from *Arabidopsis*.; bp, basepairs; CEL I, mismatch nuclease I from celery; MBN, mung bean nuclease; nt, nucleotide; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction.

REFERENCES

1. Nucleases, eds. Linn, S. M., Lloyd, R. S., and Roberts, R. J. Cold Spring Harbor Laboratory Press, 1993.
2. Shank, T. E., Rhodes, C., Rigby, P. W. J., and Berg, P. (1975) *Proc. Nat. Acad. Sci. USA*, **72**, 989-993
3. Maekawa, K., Tsunasawa, S., Dibo, G., and Sakiyama, F. (1991) *Eur. J. Biochem.* **200**, 651-661
4. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4457-4462
5. Kroeker, W. D., Kowalski, D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4463-4467
6. Sheflin, L. G. and Kowalski, D. (1984) *Nucleic Acids Res.* **12**, 7087-7104
7. Shank, T. E., Rhodes, C. Rigby, P. W. J., and Berg, P. (1975) *Proc. Nat. Acad. Sci. USA*, **72**, 989-993
8. Berk, A.J. and Sharp, P.A. (1977) *Cell* **12**, 721-732
9. Gerald, G. (1985) *Focus* 7,7-8
10. Silber, J. R., and Loeb, L. A. (1981) *Biochem. Biophys. Acta.* **656**, 256-264
11. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) *Nucleic Acids Research* **26**, 4597-4602
12. Finkelstein, E., Afek, U., Gross, E., Aharoni, N., Rosenberg, L., and Halevy, S. (1994) *International Journal of Dermatology* **33**, 116-118

13. Yang, B., Wen, X., Oleykowski, C. A., Miller, C. G., Kulinski, J. A., Besack, D. A., and Yeung, A. T. (1999) Submitted for publication.

14. Panavas, T., Pikula, A., Reid, P.D., Rubinstein, B., and Walker, E. L. (1999) *Plant Molecular Biology* **40**, 237-248

15. Abler, M.L., Thompson, D., Lebrasseur, N., van Hoof, A. and Green, P.J. (1999) Genbank accession number U90264.

16. Thelen, M.P., and Northcote, D.H. (1989) *Planta* **179**, 181-195

17. Aoyagi, S., Sugiyama, M., and Fukuda, H. (1998) *FEBS Letters* **429**, 134-138

18. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4457-4462

19. Sung, S., and Laskowski, M., Sr. (1962) *J. Biol. Chem. 237,* 506-511

20. Kowalski, D., Natale, D.A. and Eddy, M.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9464-9468

21. Katz, A. K., and Glusker, J. P. (1998) *Adv. in Mol. Structure Res.* **4**, 227-279

22. Romier, C., Dominguez, R., Lahm, A., Dahl, O., and Dietrich, S. (1998) *Proteins: structure, Function, and Genetics 32*, 414-424

FIGURE LEGENDS:

Fig. 1. **Nicking of the RF-I DNA by CEL I and mung bean nuclease.** Assays are in the presence (solid symbols) or absence (hollow symbols) of 3 mM $MgCl_2$. Panels A, C, and E are assays at pH 5.5. Panels B, D, and F are at pH 7.5.

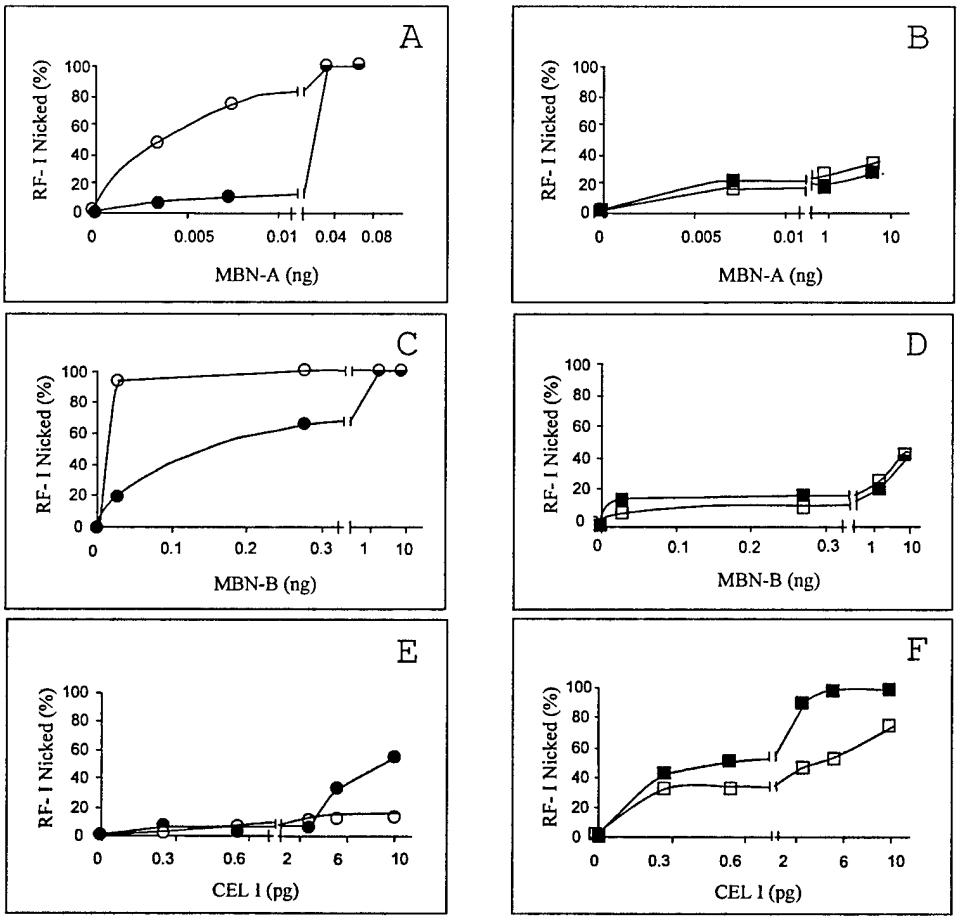
Fig. 2. **Solubilization of denatured calf-thymus DNA by CEL I and mung bean nuclease.** Assays are in the presence (solid symbols) or absence (hollow symbols) of 3 mM $MgCl_2$. Circles are assays at pH 5.5. Squares are at pH 7.5. The enzymes tested in panels A, B, and C are MBN-A, MBN-B, and CEL I, respectively. One unit of single-strand nuclease activity of CEL I equals 32 ng of homogeneous CEL I (3.1×10^4 single-strand nuclease units/mg enzyme as seen in initial kinetics up to 20 min in panel C).

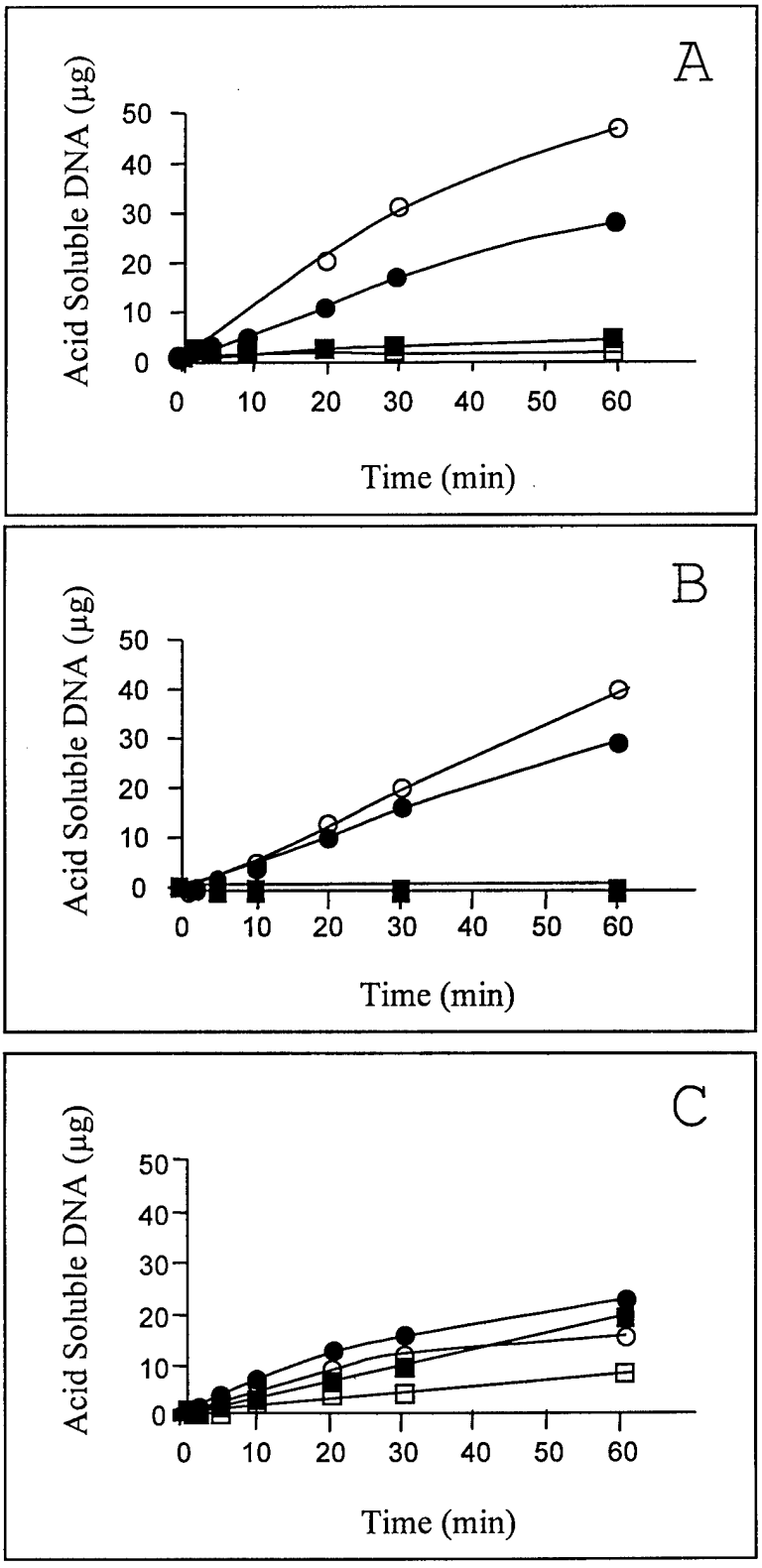
Fig. 3. **Electropherograms of mutation detection GeneScan analyses using MBN or CEL I.** Two color fluorescent heteroduplexes of PCR products of *BRCA1* gene were prepared as described in the experimental procedures. Vertical axis, relative fluorescence units; horizontal axis, DNA length in nucleotides. Panels A, D, and G, the DNA was incubated with 7 ng of MBN-A. Panels B, E, and H, the DNA was incubated with 11 ng of MBN-B. Panel C, F, and I, the DNA was incubated with 10 pg of CEL I. These reactions were performed in Buffer B with 3 mM $MgCl_2$ for 30 min at 37 °C. In panels A, B, and C, the substrate was a 387 bp heteroduplex containing a 4 nt deletion. In panels D, E, and F, the substrate was a 323 bp product containing a C→T base substitution mismatch. In panels G, H, and I, the substrate was a 402 bp heteroduplex containing a C

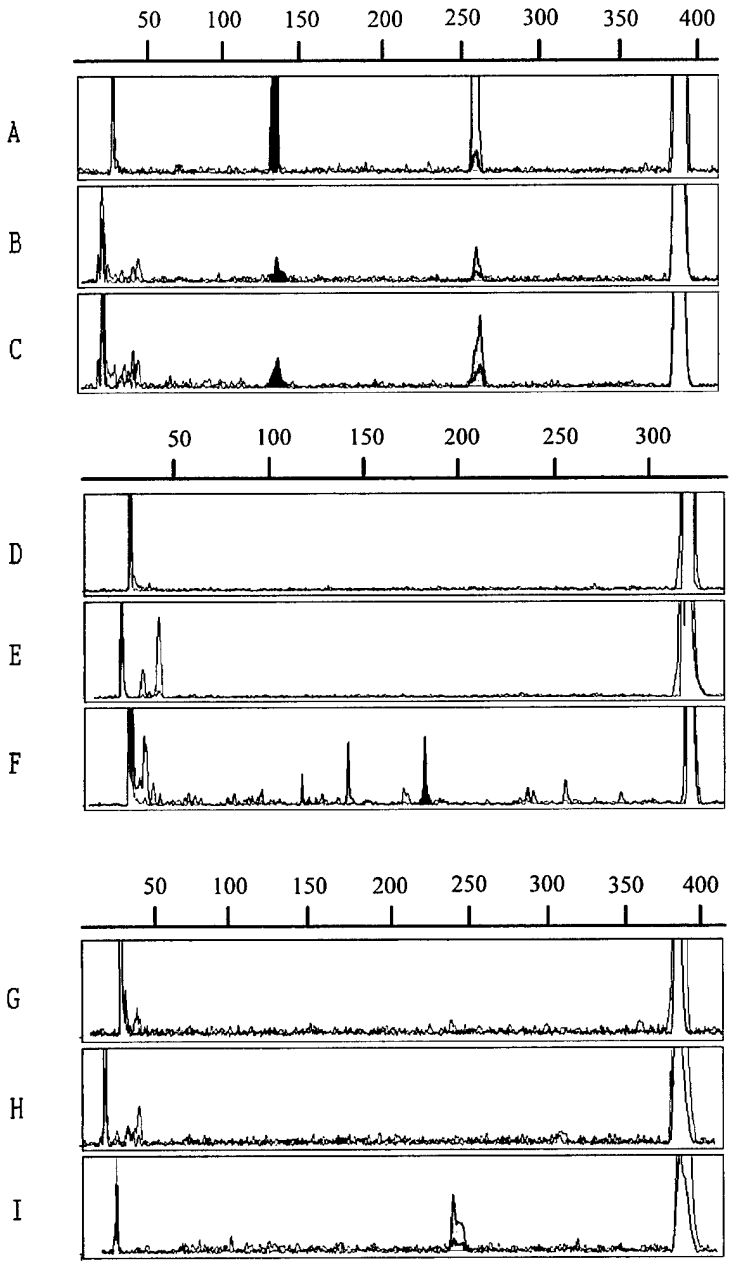
insertion in one strand. In each of panels A, B, and C the blue peak at 129 nt corresponds to cutting at the 4 base insertion on the 6-FAM-labeled strand; the green peak at 258 nt corresponds to the cutting at the 4 base insertion on the TET-labeled strand. In panels, D, E, G, and H, no mismatch-specific cutting is seen by the two MBN's. In panel F, the blue peak at 183 nt corresponds to CEL I-mismatch-specific cutting on the 6-FAM-labeled strand, and the green peak at 142 nt corresponds to the mismatch-specific cutting on the TET-labeled strand. In panel I, the green peak at 252 nt corresponds to the CEL I specific cutting at the extrahelical G on the TET-labeled strand.

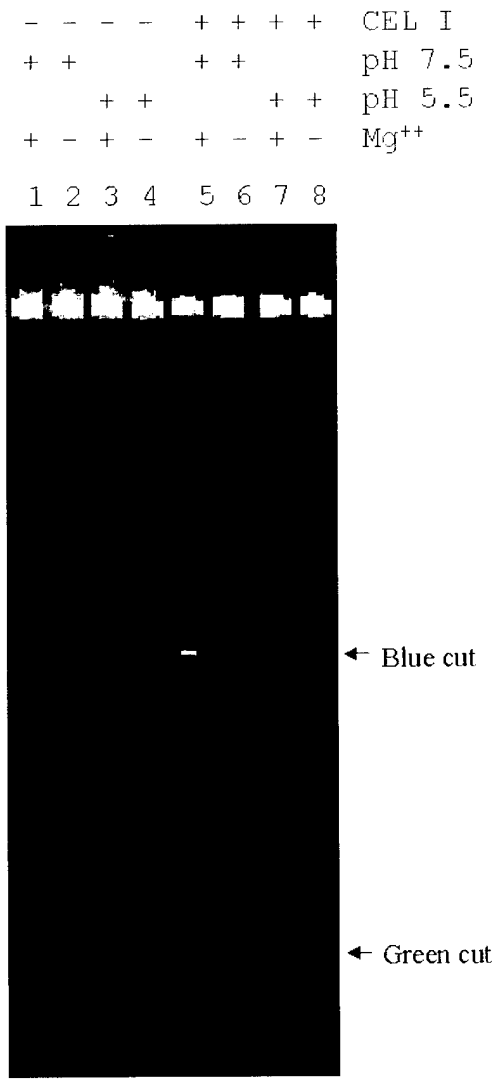
Fig. 4. **Effects of Mg^{++} and pH on CEL I mutation detection.** The picture is a gel image of mutation detection analyses on a Perkin Elmer automated DNA sequencer running the GeneScan program. The substrate is a 235 bp PCR product of the *BRCA1* gene exon 5 containing a T→G polymorphism. It is labeled at the 5' terminal with 6-FAM (Blue) in the top strand and with TET (Green) on the bottom strand. The substrates were incubated with 0.5 units of CEL I for 30 min at 45 °C and then analyzed as described in Fig. 3. In lane 5 the blue band at 156 nt corresponds to CEL I mismatch-specific cutting on the 6-FAM-labeled strand, and the green peak at 80 nt corresponds to the mismatch-specific cutting on the TET-labeled strand. The red bands in a gel image are the internal size standards in each lane.

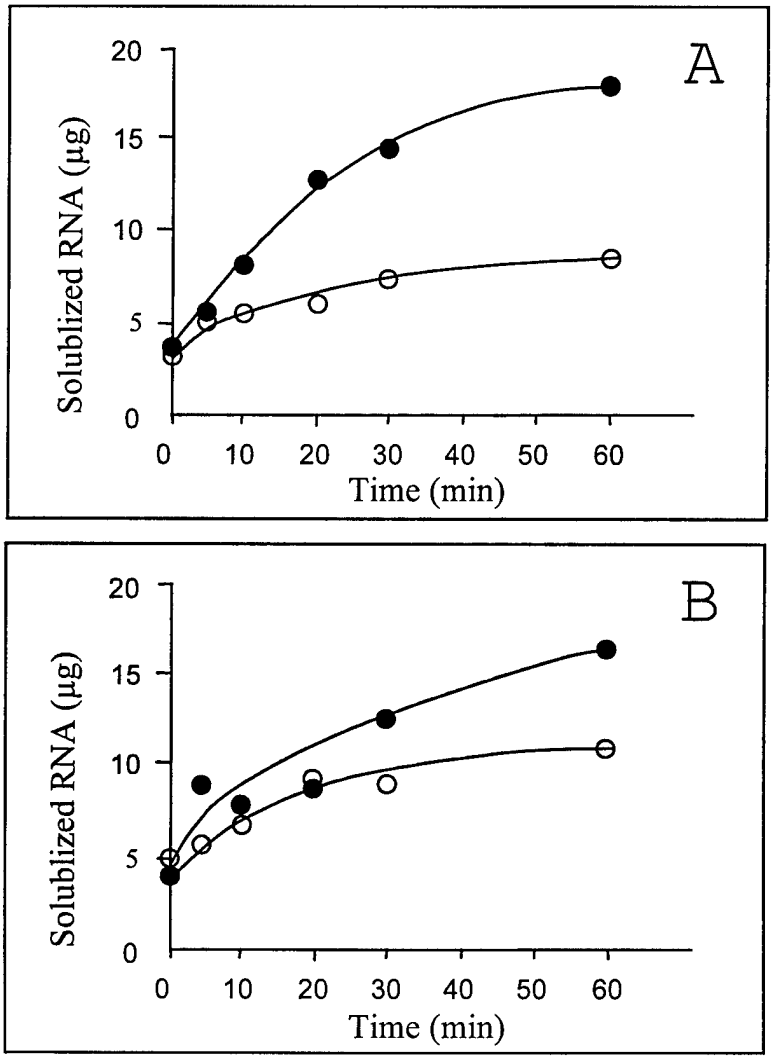
Fig. 5. **Solubilization of RNA by CEL I and mung bean nuclease**. Torula yeast RNA was incubated with 0.7 ng of MBN-1 (solid circles) or 16 ng of CEL I (hollow circles) in the presence of 3 mM $MgCl_2$ at pH 5.5 (A) and pH 7.5 (B).


A
RF- I Nicked (%)
100
80
60
40
20
0
0
0.005
0.01
0.04
0.08
MBN-A (ng)
B
RF- I Nicked (%)
100
80
60
40
20
0
0
0.005
0.01
1
10
MBN-A (ng)
C
RF- I Nicked (%)
100
80
60
40
20
0
0
0.1
0.2
0.3
1
10
MBN-B (ng)
D
RF- I Nicked (%)
100
80
60
40
20
0
0
0.1
0.2
0.3
1
10
MBN-B (ng)
E
RF- I Nicked (%)
100
80
60
40
20
0
0
0.3
0.6
2
6
10
CEL I (pg)
F
RF- I Nicked (%)
100
80
60
40
20
0
0
0.3
0.6
2
6
10
CEL I (pg)


A
Acid Soluble DNA (µg)
50
40
30
20
10
0
0
10
20
30
40
50
60
Time (min)
B
Acid Soluble DNA (µg)
50
40
30
20
10
0
0
10
20
30
40
50
60
Time (min)
C
Acid Soluble DNA (µg)
50
40
30
20
10
0
0
10
20
30
40
50
60
Time (min)


50 100 150 200 250 300 350 400
A
B
C
50 100 150 200 250 300
D
E
F
50 100 150 200 250 300 350 400
G
H
I


- - - - + + + + CEL I
+ + + + pH 7.5
+ + + + pH 5.5
+ - + - + - + - Mg++
1 2 3 4 5 6 7 8
Blue cut
Green cut


A
Solublized RNA (µg)
20
15
10
5
0
0
10
20
30
40
50
60
Time (min)
B
Solublized RNA (µg)
20
15
10
5
0
0
10
20
30
40
50
60
Time (min)

The CEL I enzymatic mutation detection assay

Joanne Kulinski[§], David Besack[§], Catherine A. Oleykowski, Andrew K. Godwin, and Anthony T. Yeung**

From the Fox Chase Cancer Center, Philadelphia, PA 19111

**to whom correspondence should be addressed:
Anthony T. Yeung
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111
Tel: (215) 728-2488
FAX: (215) 728-3647
Email: AT_Yeung@fccc.edu

§ These authors contributed equally to the work in this report

Abbreviations

The abbreviations used are: bp, basepairs; nt, nucleotide(s); PCR, polymerase chain reaction; RT, room temperature.

Abstract

CEL I, a novel endonuclease from celery, is the first nuclease known that has a high specificity for mismatches, insertions and deletions in double-stranded DNA. Our laboratory has purified this enzyme and developed it into a mutation detection assay (Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, **26**, 4597-4602, 1998). We have now optimized the CEL I mutation detection assay, and report the new parameters herein. The enzyme is found to be extremely stable. The assay functions over a wide range of buffers, salt concentrations, enzyme concentrations and incubation conditions. Two studies are reported to demonstrate the utility of the new streamlined protocol: (i) A rapid evaluation of the coding regions of the human *BRCA1* gene of 10 persons for mutations and polymorphisms, (ii) 100 people was screened for mutations and polymorphisms in a 487 bp region of *BRCA1* in a single DNA sequencing gel. The latter used multiplexing of the DNA of five people in each mutation detection reaction for one DNA sequencing lane.

Introduction

Mutations in key genes are a common basis for cancer (1-4). Many genes are polymorphic among people, and some polymorphisms can lead to biological changes (5-9). Many procedures have been devised for the detection of mutations and polymorphisms in genes. The method developed in this laboratory is called CEL I mutation detection. CEL I is an enzyme we discovered in celery (10). It is the first nuclease known to have high specificity for insertions, deletions, and base-substitution mismatches. The assay uses fluorescent or radioactive labeled nucleotides for fragment detection. Briefly, PCR is used to amplify the normal and mutant alleles of the target sequence. The PCR primers used are labeled with blue (forward primer) and green (reverse primer) fluorescent dyes. Upon denaturing and renaturing of the normal and the mutant allele in a mixture, mismatch heteroduplexes will be formed approximately 50% of the time. For each base change, two mismatches are formed (e.g.: C to T gives mismatches C/A and T/G). CEL I cuts one strand of DNA per duplex at the 3' side phosphodiester bond of a mismatch, thereby truncating one strand of the PCR product to produce a shorter DNA fragment (e.g.: blue). In another DNA molecule, CEL I cuts at the mismatch in the other DNA strand, producing a truncated DNA fragment of the second color (e.g.: green). The DNA product is analyzed on an automated DNA sequencer, Model 377 (Perkin-Elmer), and the mobility of each fragment is analyzed with Genescan software (Perkin-Elmer). The sizes of the fragments of the two colors (their sum approximately equals the length of the PCR product) independently pinpoint the location of the mismatch. When both a green peak and a blue peak are observed for one base-substitution, they represent two to four independent CEL I incision reactions in this mixture of mismatches. As such, the mutation/polymorphism detected by CEL I mutation detection method is highly reliable.

In this report, we show that buffers, salts, enzyme concentration and incubation time all have little impact on the CEL I mutation detection method. We further streamlined the procedure by removing a previous step in which the PCR product was purified by Wizard Prep (Promega). The addition of AmpliTaq DNA polymerase in the CEL I reaction has also been removed. As a test

of the efficacy of this procedure, we show (i) the new streamlined protocol was used to rapidly evaluate all the exons of the *BRCA1* gene (11) of 10 persons for mutations and polymorphisms, (ii) 100 people was screened for mutations and polymorphisms in a 487 bp region of *BRCA1* in a single DNA sequencing gel. The assay used multiplexing of the DNA of five people in each mutation detection reaction for one DNA sequencing lane.

Materials and Methods

CEL I was purified from celery to near homogeneity as previously described (10). AmpliTaq DNA polymerase and dNTPs used for PCR were from Perkin-Elmer. PCR primers were made in the Fannie E. Rippel Biotechnology Facility in our institution.

Sample Ascertainment

As part of a Fox Chase Cancer Center (FCCC) Institutional Review Board approved protocol, peripheral blood samples were obtained from consenting affected high risk family members through the Margaret Dyson/Family Risk Assessment Program (FRAP). Individuals participating in FRAP have agreed to allow their samples to be used for a wide range of research purposes including screening for mutations in candidate cancer predisposing genes, such as *BRCA1* (11). The participating individuals had previously been screened for *BRCA1* mutations by the Clinical Genetic Testing laboratory at FCCC, the results of which were later confirmed by sequencing. However, CEL I mutation detection in our current study was done in a blind fashion.

Sample Preparation

Samples were prepared for enzyme digestion by PCR amplification of genomic DNA using thirty primer pairs for the coding region of the human *BRCA1* gene [Figure 1]. The PCR reaction was performed in a 20 μL reaction volume containing 1U AmpliTaq DNA polymerase, dNTPs at 60 μM each, 75 ng of genomic DNA, and 1.0 μM primers in 1X PCR Buffer (1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl, pH 8.3 at RT). After initial denaturing step at 94°C for 4 min, the DNA was amplified through 20 cycles consisting of denaturing at 94°C for 5 sec, annealing at 65°C for 1 min decreasing by 0.5°C per cycle, and extension at 72°C for 1 min. The samples were then subjected to an additional 30 cycles consisting of denaturing at 94°C for 5 sec, annealing at 55°C for 1 min, and extension at 72°C for 1 min, with a final extension of 5 min at 72°C.

PCR samples used for the optimization study were amplified using 1.0 μM primers and purified using Wizard PCR Preps (Promega) which remove unused primers, dNTPs, and salts. The PCR samples used to test ten individuals for polymorphisms in the *BRCA1* gene and those used in the multiplexing study were amplified using 0.2 μM primers. These PCR reactions were used without further purification.

To evaluate PCR, 4 μL were loaded onto a 2% agarose gel and compared to a Precision Molecular Mass Standard (BioRad) to find the approximate concentrations of DNA in each sample, which were then adjusted to a concentration of ~10 ng/μL in dH_2O. For the multiplexing study, samples were first evaluated for concentration and then mixed in equal amounts, in pools of 5 individuals, in 1X PCR buffer.

CEL I Mutation Detection Assay
0.2U[1] of CEL I was mixed with 50 ng of DNA sample in a 10 µL reaction buffered with 3 mM $MgCl_2$, 10 mM KCl, and 20 mM Hepes, pH 7.5 at RT, unless otherwise noted. Prior to adding CEL I, samples were heated to 94°C for 1 min and slowly cooled to RT to form heteroduplexes. Samples were then incubated for 30 min at 45°C, using 1.1 µL of 10 mM o-phenanthroline to stop the reaction. Then each sample was processed through an AutoSeq G-50 column (Pharmacia) and dried using vacuum centrifugation. Each pellet was then dissolved in a mixture of 3.6 µL deionized formamide, 0.7 µL TAMRA labeled gel standard, and 0.7 µL loading dye.

While testing the effects of various buffers on the CEL I assay, the enzyme reaction buffer was adjusted appropriately. For the evaluation of the *BRCA1* gene for ten individuals and the multiplexing study, using unpurified DNA led to the presence of PCR buffer in the CEL I reaction. Because of this, multiplex reactions were buffered with 3 mM $MgCl_2$, 25 mM KCl, 20 mM Hepes, 5 mM Tris, pH 7.5 at RT.

Sample Analysis
Samples were heated at 90°C for 1 min., loaded onto a 6% polyacrylamide gel and run on an ABI 377XL DNA sequencer (Perkin-Elmer) under denaturing conditions. Each sample was analyzed using Genescan software to determine if any DNA fragments detected are the result of CEL I digestion.

Results and Discussion

Optimization of the CEL I mutation detection assay
The optimization of individual components of the assay is reported below. Because most parameter changes have no significant impact on the CEL I mutation detection protocol, less than one fold difference in signal strength, most of the conclusions are presented without showing the supporting data.

1. PCR reaction. The parameters of the PCR reaction used to amplify all target sequences of the *BRCA1* gene are the same, which often leads to high background or low signal for some exons. To minimize non-specific products that contribute to background, less DNA template is used in the PCR reaction than previously reported (1 ng/µL instead of 3.5 ng/µL) and primers are used at a concentration of 0.2-0.4 µM instead of 1 µM. With these changes, there is no change in the yield of target PCR products, but the background is often lower. This helps eliminate the need for purification of the PCR product. In addition, primers are designed such that the first 2 or 3 bases at the 5' end are G or C. This is reflected in the newly designed primers for exon 11 in Figure 1, but not necessarily true of other exons. PCR products made with these primers are more resistant to CEL I terminal digestion and help to preserve the mismatch incision signals of the two colors. PCR product lengths are kept at or below ~500 bp whenever possible because it is easier to handle in gel analysis.

[1] One unit of CEL I is defined as 1/1000 of one unit of single-stranded nuclease activity. One unit of single-stranded DNase activity is defined as the amount of enzyme that produces 1 µg of acid-soluble material at pH 5.5 in 1 min. at 37°C in the absence of magnesium when purified sheared single-stranded calf thymus DNA is used as the substrate.

2. Enzyme Stability. CEL I is stable during storage as a 10X working solution (in 10X reaction buffer containing 50% glycerol) at –70°C or –20°C for at least one year. It can be stored at 4°C for several weeks as well as at RT for a few days, with no noticeable change in mutation detection ability. It is stable to at least 20 cycles of freezing and thawing.

3. Incubation conditions. For the purpose of mutation detection, the incubation with CEL I can be from 0.5 Hrs to 5 Hrs [Figure 2]. The assay can also accommodate a four fold variation in enzyme concentration as well as a wide range of DNA concentrations with little change in the results. There is sufficient latitude in the assay and in the software analysis such that a low signal with low background can be analyzed as easily as a strong signal with high background. However, it is important to note that doubling incubation time is not equal to doubling enzyme concentrations. This is consistent with the fact that more than one enzyme-substrate interaction occurs in the mutation detection reaction. Previously, AmpliTaq DNA polymerase was found to stimulate CEL I in assays using ^{32}P labeled oligonucleotides (10). This stimulation does not occur in the case of PCR products 5'-labeled with a fluorescent group due to reasons that are unclear at this point. As a result, AmpliTaq is omitted in the current assay configuration.

4. Buffers. The CEL I mutation detection assay is not sensitive to moderate changes in buffer conditions. The original buffer was 20 mM Tris•HCl, 10 mM $MgCl_2$, 25 mM KCl. Hepes has replaced Tris in the assay due to its pK near 7.5, the pH of the reaction buffer. The assay performs well from pH 7 to 9, but not below 6. Potassium glutamate from 1-100 mM has no noticeable effect on the assay, as well as 1-30 mM Triethylamine acetate. Phosphate $\leq$ 45 mM slows down the assay, but does not inhibit it completely.

5. Metals. For divalent metal requirement, mutation detection assay is optimal for 3-10 mM Mg^{2+} in the incubation buffer. $MgSO_4$ and $MgCl_2$ worked identically. Calcium can replace magnesium completely in the assay. When cobalt or manganese are used to replace magnesium, non-mismatch cutting dominates over mismatch cutting, resulting in rapid degradation of PCR product. For monovalent cations, Li^+, K^+, Na^+, and Cs^+ are comparable in the assay, but not a necessary component. Varying the salt concentration in the buffer from 0 to 30 mM has no apparent effect on the assay.

6. Sample preparation. After CEL I reaction, samples are processed through an AutoSeq G-50 column (Pharmacia) to remove buffer components after the CEL I reaction is complete. Without this step, fragments less than ~100 bases do not migrate properly in the gel, making analysis in this region difficult. Since two cuts are formed for every mismatch, any small fragment will have a corresponding large fragment "partner" that will in theory independently give mismatch location, but this cannot be depended upon. Sometimes having signal of only one color is not convincing, especially in areas of high background. The benefits of this step may be an important consideration when there is no prior knowledge of what mutation is to be expected in a sample.

Testing the CEL I assay without purifying the PCR product

Study #1: Results of evaluation of entire *BRCA1* gene without PCR product purification. The coding region of the *BRCA1* gene is over 5 Kbp, divided into 24 exons. Moreover, exon 11 is large and needs to be divided into several smaller fragments for mutation detection. In this

experiment, twenty working days were spent on PCR amplification of samples, agarose gel analysis of the PCR, CEL I digestion reactions, and gel analysis of the digestion products. Five working days were spent on software analysis of the data to call the mutations and polymorphisms.

One is able to perform PCR using 0.2 μM primers for the entire gene using the 30 primer pairs given in Figure 1. The entire coding region of *BRCA1* was evaluated using one PCR product for each exon and 9 PCR products for exon 11. Primers for exon 11 were designed to produce PCR products ≤ about 550 bp in length with a 100 bp overlap.

Out of the 10 individuals, there were 37 polymorphisms in the coding regions and 10 in an intron region [Figure 3]. CEL I produced mismatch specific cuts for all 47 polymorphisms with no false positives. However, some experience and care is important when a large volume of data is being analyzed. This study shows the feasibility of mutation detection by CEL I without purification of the PCR products. The improvement is a considerable benefit since purification by Wizard Prep is time consuming for a study with many samples involved (30 primers pairs, 10 individuals is 300 samples).

Study #2: Evaluation of 100 individuals for a 487 bp PCR product on one gel by multiplexing PCR products.

Five unpurified PCR products were mixed per CEL I reaction, and analyzed in one gel lane (100 individuals, 20 reactions, one Genescan gel). The target was section 4 of exon 11 of the *BRCA1* gene. No mutations were found in any individual for this region, although 3 different polymorphisms were found (#'s 2, 3, 4 of Figure 3). Representative electropherograms for this data are in Figure 4. This test provides evidence that it is practical to detect mismatches in mixtures of 5 samples, although the determination of which one of the 5 samples is positive will require analysis of one sample at a time. Our example sequence is a highly polymorphic region of the *BRCA1* gene and it contains multiple polymorphisms in all the DNA pools. The frequency of the polymorphisms, as determined one person at a time in study #1, is 3/10. 5/10, and 4/10, for polymorphisms #2, #3, and #4, respectively. With such frequencies, our sample of 20 DNA pools will contain several samples at a polymorphism allele frequency of 1/5. If the mutation or polymorphism being determined is relatively rare, this multiplexed analysis will allow its discovery about 5 times faster than analyzing one person's DNA at a time.

This test demonstrates the ability of CEL I to detect more than one polymorphism in a PCR product, a characteristic not present in other mutation detection assays. It is also shown that CEL I can detect 2 polymorphisms that are 5 nt apart from one another (i.e. polymorphisms 2201 and 2196, panels D, E of Figure 4).

In summary, we have presented the optimized conditions for the CEL I mutation detection assay. As seen for the two studies reported herein, this assay is accurate, reliable, and fairly straight forward to set up.

Current CEL I Mutation Detection protocol and conditions:
Primer Concentration for PCR: 0.2-0.4 μM each
Substrate: ~50 ng PCR amplified DNA per reaction
Purification of PCR Reaction: Wizard Prep purification of the PCR products is used sometimes, but not necessary.
Enzyme: 0.4 U CEL I per reaction
Buffer: 3 mM $MgCl_2$, 10 mM KCl, 20 mM Hepes pH 7.5 at room temperature.
Incubation: 1 hour at 45°C, stopped with 1 mM o-phenanthroline.

Acknowledgements

We thank C. Glenn Miller and Bob Muhlhauser for the synthesis of the fluorescent PCR primers.

Footnotes

† This work was supported, in part, by grants NIH CA71426 and U.S. Army DMAD17-97-1-7286 to ATY, by institutional grants from the National Institute of Health to Fox Chase Cancer Center (CA06927, RR05539), and an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. Castilla, L. H., Couch, F. J., Erdos, M. R., Hoskins, K. F., Calzone, K., Garber, J. E., Boyd, J., Lubin, M. B., Deshano, M. l., Brody, L. C., Collins, F., and Weber, B. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nature Genet.* 8, 387-391, 1994.
2. Friedman, L. S., Ostermeyer, E. A., Csilla, I. S., Dowd, P., Lynch, E. D., Rowell, S. E., and King, M. C. Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genet.* 8, 399-404, 1994.
3. Nicolaides, N. C., Papadopoulos, N., Liu, B., Wel, Y.-F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltin, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Dunlop, M. G., Hamilton, S. R., Peterson, G. M., Chapelle, A. , Vogelstein, B., and Kinzler, K. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 371, 75 1994.
4. Bronner, C. E., Baker, S. M., Morrison, P.T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R. J., Godwin, A., Ward, D. C., Nordenskjold, M., Fishel, R., Kolodner, R., and Liskay, R. M. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 368, 258-261, 1994.
5. Racchi, M., Watzke, H.H., High, K.A., Lively, M.O. (1993) Human coagulation factor X deficiency caused by a mutant signal peptide that blocks cleavage by signal peptidase but not targeting and translocation to the endoplasmic reticulum. *J Biol Chem* 268:5735-5740
6. Brennan, S.O., Myles, T., Peach, R.J., Donaldson, D., George, P.M. (1990) Albumin Redhill (-1 Arg, 320 Ala-Thr): a glycoprotein variant of human serum albumin whose precursor has an aberrant signal peptidase cleavage site. *Proc Natl Acad Sci U S A* 87:26-30
7. Ito, M., Oiso, Y., Murase, T., Kondo, K., Saito, H., Chinzei, T., Racchi, M., Lively, M.O. (1993) Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes insipidus. *J. Clin. Invest.* 91:2565-2571
8. Daly, M., Bruce, D., Perry, D.J., Price, J., Harper, P.L., O'Meara, A., Carrell, R.W. (1990) Antithrombin Dublin (-3 Val----Glu): an N-terminal variant which has an aberrant signal peptidase cleavage site. *FEBS Letters* 273:87-90
9. Thompson, S. J., Robinson C., and Mant., A. Dual signal peptides mediate the signal recognition particle/Sec-independent insertion of a thylakoid membrane polyprotein, PsbY. *J. Biol. Chem.* 1999; 274, 4059-66
10. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, 26, 4597-4602.
11. Miki, Y., Swenson, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Nennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonnin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Golgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994) . A strong candidate gene for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 266, 66-71.

Figure 4

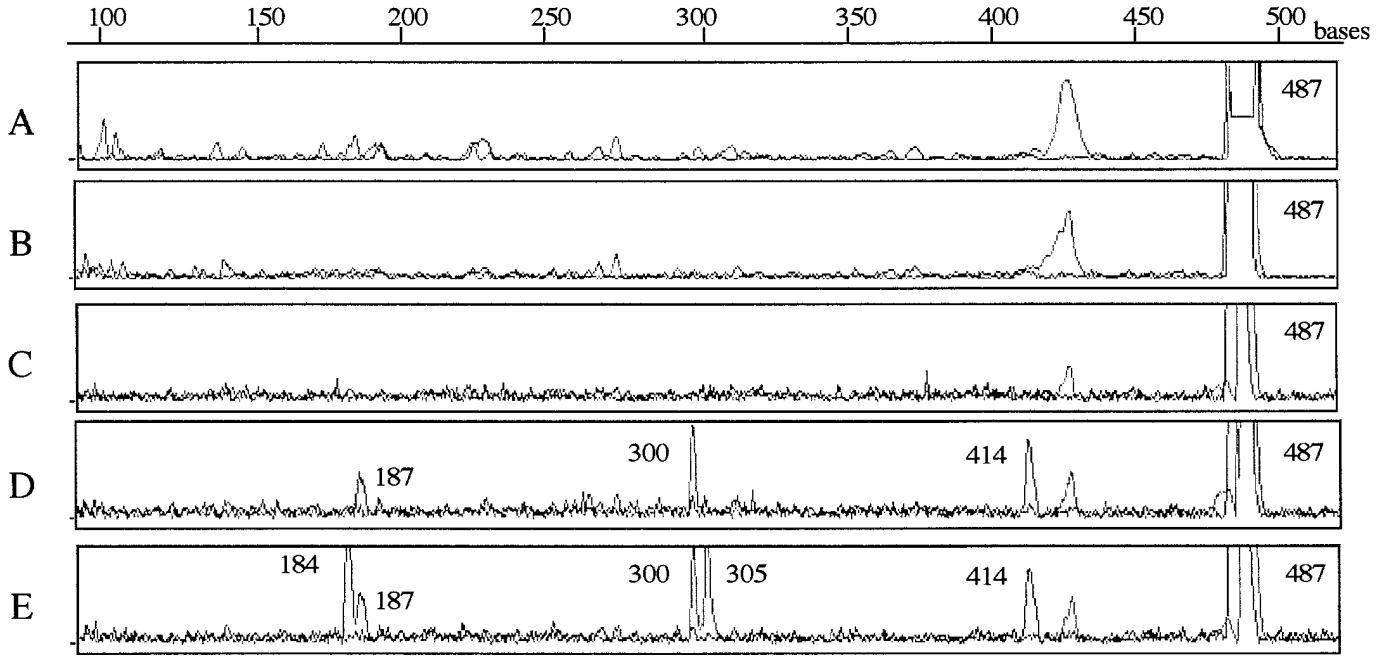


Figure 4: **Representative data of the CEL I mutation detection analyses of multiplex samples.** Electropherograms of gel lanes of multiplex samples of exon 11 section 4. Horizontal axis is DNA fragment length and vertical axis is the relative fluorescence, reflecting the amount of fragment. Each panel shows ~10 ng of fluorescent labeled PCR amplified DNA. Panel A was not reacted with CEL I, panels B-E were reacted with 0.2U CEL I for 30 min. Panel B contains control DNA with no polymorphisms. Of the 20 pools of 5 individuals each (100 individuals total), 1 contained no polymorphisms (C), 8 contained polymorphisms at nt 2201 and 2430 (D), and 11 contained polymorphisms at nt 2196, 2201, and 2430 (E). In each panel, the green peak at about 425 bases is a PCR band, not the result of CEL I digestion, evidenced by its presence in panel A. The fragments of length 187 and 300 are produced from the same polymorphism (nt 2201), as are the fragments of length 184 and 305 (nt 2196). Alternatively, the green peak corresponding to the polymorphism at base 414 (nt 2430) is not seen because it is short enough to get lost among the small DNA fragments at the bottom of the gel.

Polymorphisms of The Human Steroid Sulfatase Gene

David Besack, Joanne Kulinski, Catherine A. Oleykowski, Rebecca Raftogianis‡, Andrew K. Godwin‡, and Anthony T. Yeung**

Running Title: **ARSC polymorphism**

From the Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111

‡ The Medical Sciences Division, Fox Chase Cancer Center, Philadelphia, PA 19111

** To whom correspondence should be sent: Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111. Tel: (215) 728-2488. FAX: (215) 728-3647. Email: AT_Yeung@fccc.edu

Abbreviations

ARSC, bp, basepair; human arylsulfatase C isozyme S (steroid sulfatase); m/p, mutation/polymorphism; nt, nucleotide; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism

Summary

The humans steroid sulfatase C (ARSC) enzyme is important to the proper balance of active ARSC enzyme in the body. This study reports the result of a screen of 100 random individuals for potential polymorphisms in the coding regions of this gene. The CEL I mutation detection method is used in this study in which the polymorphisms were ascertained twice by using two protocols that gave the same result. The first protocol uses one pair of fluorescently labeled PCR primers for each exon. The second protocol uses only one pair of universal fluorescent PCR primers in the screening of all the ARSC exons. Our results indicated that the ARSC coding region is remarkably low in polymorphism. The only mutation/polymorphism found in the coding regions among these 100 individuals is a Met to Ile missense change at the sixth amino acid from the N-terminal of the nascent protein of one individual.

Introduction

Sulfation and desulfation are important reactions in the metabolism of many steroid hormones (1, 2). Estrone, estradiol and dehydro-epiandrosterone (DHEA) circulate predominantly in the sulfated form and as such are not biologically active (i.e., do not bind target receptors). Furthermore, the sulfated forms of many steroid hormones exhibit half-lives up to ten-fold higher than the desulfated form. Biological "cycling" of sulfated/desulfated steroid hormones has been demonstrated. The sulfated moiety represents a readily accessible, yet biologically inactive, "storage" form for many steroid hormones whereby hydrolysis of the sulfate group (desulfation) regenerates the biologically active steroid. These observations suggest that sulfation and desulfation represent important reactions in the regulation of the biological activity of steroid hormones, and this regulatory system has become a target for chemotherapy of steroid hormone dependent tumors. ARSC, also known as steroid sulfatase (STS), catalyzes the desulfation of estrone-, 17β-estradiol-, and DHEA sulfate. As a first step to investigate whether functionally significant genetic polymorphisms occur within ARSC, we analyzed the ARSC structural gene of 100 persons for the possible presence of polymorphisms.

Mutation and polymorphism identification was performed using a new CEL I endonuclease assay (3). CEL I, isolated from celery, is the first eukaryotic nuclease known that cleaves DNA with high specificity at sites of base-substitution mismatch and DNA distortion. DNA is cut at the phosphodiester bond 3' of the mismatch in one of the two strands of a heteroduplex. The mutation detection assay is based on fluorescence detection, analysis is performed on an automated DNA sequencer, and the data is easily analyzed with the Genescan software of Perkin-Elmer. The assay is reliable, and normally does not produce false positives or false negatives. In this report, we also illustrate an alternate approach, using one pair of universal fluorescent primers, along with unlabelled primers, to screen all the exons of the ARSC gene. Our result shows that the ARSC is remarkably limited in polymorphism at the nucleotide level.

Subjects and Methods

Subjects

Sample population consisted of purified DNA collected from 100 individuals that were tested for mutations in *BRCA1* under Fox Chase Cancer Center's Family Risk Assessment Program (FRAP). This program exists to screen those who are designated as high risk for breast cancer due to family history for mutations in *BRCA1* gene (4). Before DNA samples from these individuals were used for this study, they have already been determined to have no mutations in *BRCA1*. Subjects consisted of 92 females, 6 males, and 2 unknown.

PCR

Genomic DNA was used as template to PCR amplify all 10 exons and promoter region. A section of 3' UTR was also amplified based on information collected from the Genetic Annotation Initiative[1]. Reactions were done using 1 U AmpliTaq Gold DNA polymerase (Perkin-Elmer), 0.4 μM primers, 200 μM dNTPs, and 5% DMSO in a 20 μL reaction buffered with 2 mM $MgCl_2$, 50 mM KCl, and 10 mM Tris-HCl pH 8.3. The PCR reaction began with an initial 10 min pre-incubation at 94 °C, then proceeded for 35 amplification cycles denaturing at 94 °C for 10 sec, annealing at 55 °C for 30 sec, and elongating at 72 °C for 45 sec. Fluorescent primer sets were designed to include each exon and at least 40 nt on either side [Table 1]. Primers were labeled with either 6-FAM (forward) or TET (reverse). Samples were PCR amplified and tested in mixtures of two so that no samples from males were reacted by themselves (at least 2 alleles per sample necessary to form a heteroduplex substrate for CEL I) and to reduce the overall number of reactions.

Nested PCR for the universal fluorescent primer method

The universal fluorescent PCR primer method, using double nested PCR primers, is shown in schematic form [Figure 1]. The first round of PCR amplification is performed with unlabeled primers containing a common 5' 12 nt overhang. All forward primers contained the sequence 5' TGTGCGGTCCTC 3' and all reverse primers contained the sequence 5' TTGATCCTACAA 3'. A second round of PCR was then carried out using a single pair of universal fluorescent primers for all products. These were forward primer 5' *6-FAM* GCCAGAGTTGTGCGGTCCTC 3' and reverse primer 5' *TET* GCCCGACTTTGATCCTACAA 3'. The 3' 12 nt of these primers contain the same sequence as the respective 12 nt overhang of the unlabeled primer. The product of the first round of PCR was diluted 1:50 in H_2O and used as the template for the second round of PCR. First round PCR conditions were the same as for the fluorescent method except primers were used at 0.2 μM. Second round PCR conditions were the same as the first, except about 0.2 ng of template was used and only 15 cycles of amplification were used. Again, samples were tested in mixtures of two.

[1] CGAP-GAI Home Page (http://lpg.nci.nih.gov/GAI). This site collects GenBank entries and compares EST sequences to find sites of frequent polymorphisms.

Enzyme Reaction

0.4 U of CEL I[2] was mixed with 5 μL of PCR product in a total 10 μL reaction buffered with 3 mM $MgCl_2$, 10 mM KCl, and 20 mM HEPES pH 7.5. Reactions were incubated for 1 hour at 45 °C and stopped with 1.1 μL 10 mM o-phenanthroline. Sample were then processed through a 600 μL column containing Sephadex G-50 resin (Pharmacia) and dried by vacuum centrifugation. The DNA pellet was then dissolved in a mixture of 3.6 μL deionized formamide, 0.7 μL TAMRA labeled gel standard, and 0.7 μL loading dye.

Genescan Analysis

Samples were heated at 94 °C for one minute then loaded onto a 6% acrylamide gel run on an ABI 377XL DNA sequencer (Perkin-Elmer) under denaturing conditions. Each gel contained samples for analysis plus two samples not reacted with CEL I to use as negative controls. The Genescan software records fluorescent data collected during gel electrophoresis and creates a computer file containing a gel image with blue (6-FAM) and green (TET) PCR bands and red (TAMRA) internal standard bands. Reacted samples are then compared to unreacted samples and to each other to determine if any bands are the result of CEL I activity. If any samples were found to have bands produced by CEL I, the two individuals that make up that sample were tested again in separate reactions. This follow up testing included a post-PCR DNA purification step using Wizard PCR Preps (Promega), which removes PCR primers from the reaction and minimizes background bands seen in the gel image.

Results

Mutation/Polymorphism Identification

A missense mutation/polymorphism (m/p), G to C, was found in exon 2, nt 7 of one individual [Figure 2]. This corresponds to amino acid 6 of the ARSC protein producing a missense change from methionine to isoleucine. Another m/p (G to A) was found in an intronic region, 37 nt after the last nt of exon 9. This polymorphism was found in two individuals. These m/p's were initially found using both mutation detection protocols and were later confirmed by sequencing.

Identification of SNPs

No SNPs were found in the coding region or the promoter region. Only one SNP was found (A or G) in the 3' untranslated region, 3,922 nt past the stop codon. 38 individuals were found to be heterozygous for this polymorphism. Others were homozygous for either allele (if female), or contained only one allele (if male).

[2] One unit of CEL I is defined as 1/1000 of one unit of single-strand nuclease activity. One unit of single-strand DNase activity is defined as the amount of enzyme that produces 1 μg of acid-soluble material at pH 5.5 in 1 min at 37 °C in the absence of magnesium when purified sheared single-stranded calf thymus DNA is used as the substrate.

Discussion

Low frequency of polymorphisms

The near absence of polymorphisms in the ARSC coding region is unexpected, given that the 100 individuals examined are unrelated. We are confident that we have not missed any significant numbers of polymorphisms that may have been present given our experience with the CEL I mutation detection assay, and the correct identification of the one known polymorphism in the 3' UTR in 38 of the samples. Moreover, CEL I mutation detection of the ARSC gene was performed twice, first with individual fluorescent PCR primers for each exon, and next with the universal fluorescent PCR primer approach. The results were identical with both methods, giving us confidence that no polymorphisms have been missed in this analysis.

Signal Peptides

The missense m/p at Met6-Ile6 is a change in the sequence of the 21 amino acid long signal peptide of the ARSC protein. Because the signal peptide will ultimately be absent in the mature protein, the m/p has no effect on the enzymology of the mature ARSC protein. Some signal peptide mutations are known to have an effect on the maturation of the protein (5-9). However, given the conservative change of Met to Ile, and the distance from the cleavage site, one might not expect this m/p to affect the cleavage of the ARSC nascent protein. Future experiments will determine whether this single known polymorphism has an impact on the expression of ARSC.

CEL I mutation detection method

The CEL I mutation detection method was found to be highly reliable and expedient for this study, having screen the 10 exons of the ARSC gene of 100 individuals twice in 40 working days. An additional 20 working days were spent designing primers and testing the new method for reliability before proceeding with mass screening. In future screening, only one of the two PCR approaches is needed to screen a new gene for the first time. The rate-limiting step of this approach was the availability of instrument time for the automated DNA sequencer. We would project that if the availability of the automated DNA sequencer is not an issue, and assay conditions (including primer pairs) are set, then it should be possible for a single operator to screen 100 individuals for m/p's in a 2 Kbp gene of approximately 10 exons in about 30 working days.

The universal fluorescent primer method

This method was developed to facilitate a laboratory in starting to use the CEL I mutation detection method. All that is involved is the synthesis of unlabeled primers that adds the two universal 12 nt handles to their respective 5' ends. We showed in this study that the universal fluorescent primer approach is a viable alternative to having two fluorescent primers synthesized for each exon.

In summary, we have screened the ARSC structural gene of 100 random individuals and have found that this gene is remarkably low in polymorphisms. The CEL I mutation detection method is instrumental in accomplishing this screen expediently and reliably, and should be useful for others who wish to screen for SNPs in their genes of interest.

Acknowledgements

This work was supported, in part, by grants NIH CA71426 and U.S. Army DMAD17-97-1-7286 to ATY, by institutional grants from the National Institute of Health to Fox Chase Cancer Center (CA06927, RR05539), and an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Drs. Xxxxx and xxxxx for critical reading of this manuscript.

Electronic-Database Information

1. GenBank (http://www.ncbi.nlm.nih.gov/irx/genbank/query_form.html)
2. Genetic Annotation Initiative (http://lpg.nci.nih.gov/GAI)

References

1. Ballabio, A. and Shapiro, L.J. (1995) in *The Metabolic and Molecular Basis of Inherited Disease* (Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D., eds) pp2999-3022, "Steroid Sulfatase Deficiency and X-linked Ichthyosis". McGraw-Hill Inc., New York.
2. Alperin, E. S., and shapiro, L. J. (1997) Characterization of point mutations in patients with X-linked ichthyosis. Effects on the structure and function of the steroid sulfatase protein. *J. Biol. Chem.* **272**: 20756-20763
3. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research* **26**: 4597-4602
4. Miki, Y., Swenson, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Nennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonnin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Golgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994) A strong candidate gene for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**:66-71.
5. Racchi, M., Watzke, H.H., High, K.A., Lively, M.O. (1993) Human coagulation factor X deficiency caused by a mutant signal peptide that blocks cleavage by signal peptidase but not targeting and translocation to the endoplasmic reticulum. *J Biol Chem* **268**:5735-5740
6. Brennan, S.O., Myles, T., Peach, R.J., Donaldson, D., George, P.M. (1990) Albumin Redhill (-1 Arg, 320 Ala-Thr): a glycoprotein variant of human serum albumin whose precursor has an aberrant signal peptidase cleavage site. *Proc Natl Acad Sci U S A* **87**:26-30
7. Ito, M., Oiso, Y., Murase, T., Kondo, K., Saito, H., Chinzei, T., Racchi, M., Lively, M.O. (1993) Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes insipidus. *J. Clin. Invest.* **91**:2565-2571
8. Daly, M., Bruce, D., Perry, D.J., Price, J., Harper, P.L., O'Meara, A., Carrell, R.W. (1990) Antithrombin Dublin (-3 Val----Glu): an N-terminal variant which has an aberrant signal peptidase cleavage site. *FEBS Letters* **273**:87-90
9. Thompson, S. J., Robinson C., and Mant., A. (1999) Dual signal peptides mediate the signal recognition particle/Sec-independent insertion of a thylakoid membrane polyprotein, PsbY. *J. Biol. Chem.* **274**: 4059-4066

Table 1

		Fluorescent Primers	Cold Primers
Promoter	F	GCC TGT TCC TGC TGT AA	AAC CGC TTG GGT ATC A
	R	CAG GCC AAT CCT ACT CAA	CTG CAC CAG TGG GA
Exon 1	F	GTC TGC ATT TAT CTT TGA CAC A	AGT GTC TCC GCC TCA
	R	GGA GAA TAT GCA ACA CAA GC	AGG TAG CTG CTG TGA ACA
Exon 2	F	GTC TCA AGC TGA CAT CCT TCA	TCA AGC TGA CAT CCT TCA
	R	GGG GAC TGT TGC CTA TGA	GGG ACT GTT GCC TAT GA
Exon 3, 4	F	GCC TGG TGA CAG AGT GAG A	CAG CCT GGT GAC AGA
	R	CCA GGA AAG TCA TCC CTA AGA	CTC CCA CTC TTT TGC TAA
Exon 5	F	GGA TTG GAA TCA GGG TGT TTA	GGG TGT TTA TTG GGA CTG
	R	CCA CGA GAA ATA ACC CAG AA	GCA GCA TCA GAG GAC AAG
Exon 6	F	GGT GGC AGA CAT ACT TAA CA	GTG GCA GAC ATA CTT AAC A
	R	GGA GGC AAA GAC TTA GCA	CAG CTT TCT AAG CAC TCA
Exon 7	F	CCC ACT GAG TAG GGC AA	CAC TGA GTA GGG CAA CCA
	R	CGG ATG AGC TGA GAG G	AGT GAC CAG CGG ATG A
Exon 8	F	GGA TTG AAA TCT CCC TTG	CTC CCT TGT TGC CTC TTA
	R	GCT GTG AAA TCA GAG CTC A	GCA TAC TGG GCT GTG AA
Exon 9	F	GGA CAT TTG AGA ACA CAG GA	AGC TCC CTC ATG CTC TTA
	R	GCC ACC TTT TTA CCC TTT AG	GTT GGC CTC CAT TGA
Exon 10	F	CCG CAT CAC TTT TTC A	CCT AAT GCC GTT TCC A
	R	CTC TCA GGC GTG TTT GTA	CTC TCA GGC GTG TTT GTA
3' UTR	F	CCC CAT ATC TGT TCA ACC	CCC CAT ATC TGT TCA ACC
	R	GGC AGT GGA TGG AAG A	GGC AGT GGA TGG AAG A

Table 1. PCR primers used to screen the ARSC gene.
Each set of forward (F) and reverse (R) primers code for a sequence spanning the region named plus about 40 bp or more on both sides. Fluorescent primers were labeled with 6-FAM (forward) or TET (reverse). These tags produce PCR fragments that are blue and green, respectively, when viewed using Genescan software. Cold (unlabeled) primers contained a common sequence of 12 bases on the 5' end for use with a nested PCR method using a common fluorescent primer set. Forward primers contained the sequence 5'-TGTGCGGTCCTC-3' and reverse primers contained the sequence 5'-TTGATCCTACAA-3'.

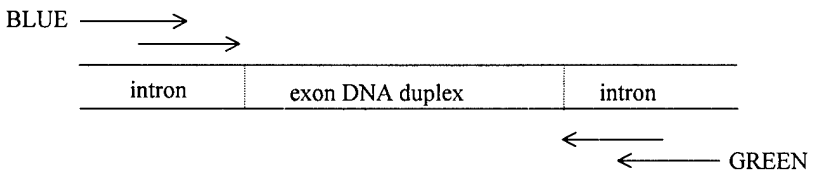


Figure 1. Double-nested PCR method for using universal fluorescent PCR primers
First round of PCR is done using unlabeled internal primers (internal arrows), the product of which is used as template for the second round of PCR using a common pair of fluorescent primers for all products (external arrows). The forward fluorescent primer contains a common sequence to the forward internal primer. The reverse fluorescent primer contains another sequence in common with the reverse internal primer. Thus, by adding these two common sequences to the 5' end of any pair of internal primers, one pair of fluorescent primers suffice for screening the entire ARSC gene.

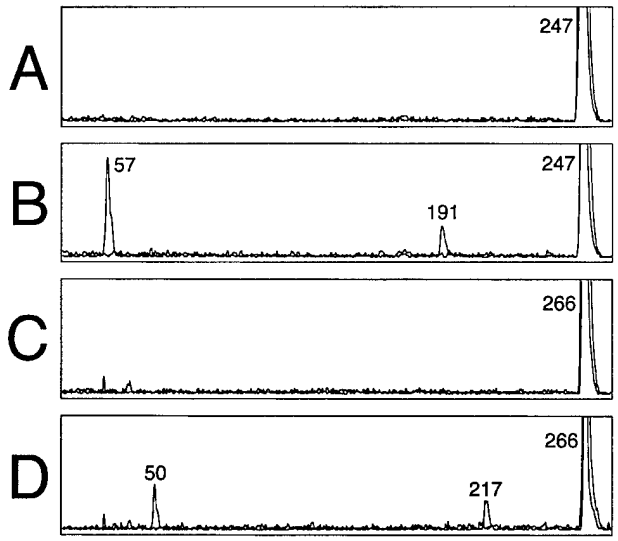


Figure 2. Electropherograms of 2 regions containing mutations.
These electropherograms represent four gel lanes, where horizontal axis is size of DNA fragment and vertical axis is intensity (amount) of fragment. Each lane in this display normally presents two chromatograms of two colors. Two representative samples of exon 2 (A, B) and exon 9 (C, D) are given, with PCR amplified DNA of length 247 bp and 266 bp, respectively. Lanes A and C contain DNA with no mismatch, while B contains a single mismatch 191 bases from the 5' end of the forward strand, producing a blue band by cutting at 191 and a green band by cutting at 57. Similarly, lane D contains a single mismatch producing a blue band by cutting at 217 and a green band by cutting at 50. Note the reduction in peak area of the PCR band in lanes containing a mismatch.

US005869245A

United States Patent [19]

Yeung

[11] Patent Number: 5,869,245

[45] Date of Patent: Feb. 9, 1999

[54] **MISMATCH ENDONUCLEASE AND ITS USE IN IDENTIFYING MUTATIONS IN TARGETED POLYNUCLEOTIDE STRANDS**

[75] Inventor: **Anthony T. Yeung**, Havertown, Pa.

[73] Assignee: **Fox Chase Cancer Center**, Philadelphia, Pa.

[21] Appl. No.: **658,322**

[22] Filed: **Jun. 5, 1996**

[51] **Int. Cl.**[6] **C12Q 1/68**; C12N 19/16; C12N 19/14

[52] **U.S. Cl.** **435/6**; 435/196; 435/195; 435/91.2

[58] **Field of Search** 435/6, 91.2, 183, 435/195, 196; 530/350

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,459,039	10/1995	Modrich et al.	435/6
5,571,676	11/1996	Shuber	435/6

FOREIGN PATENT DOCUMENTS

WO9302216	7/1992	WIPO .	
WO9320233	10/1993	WIPO .	
WO9529251	11/1995	WIPO	C12P 19/34

OTHER PUBLICATIONS

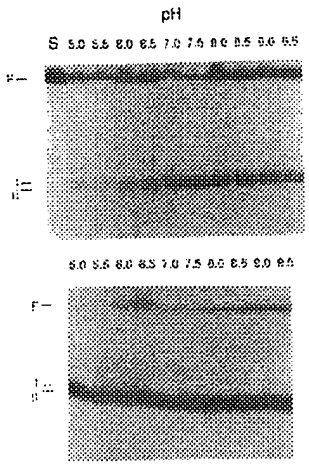
Modrich, P., 1994 Science 266, 1959–1960.
Su, et al., 1988 J. Biol. Chem. 263, 5057–5061.
Finkelstein, E., et al. 1994 International Journal of Dermatology 33, 116–118.
Smith, P.K., et al. 1985 Analytical Chemistry 150, 76–85.
Jay et al. 1990 Analytical Biochem. 185, 324–330.
Yeung, A.T., et al., 1983 Proc. Nat. Acad Sci. USA, 80, 6157–6161.
Yeung, A.T., et al., 1988 Nucleic Acids Res, 16, 4539–4554.
Lyamichev, V., et al., 1993 Science 260, 778–783.
Ramotar, D., et al., 1987 J. Biol. Chem 262, 425–31.
Chow, T.Y.K., et al., 1987 J. Biol. Chem 262, 17659–17667.
Wyen, N.V., et al., 1971 Biochem Biophys Acta. 232, 472–483.
Brown, P.H., et al., 1987 Eur. J. Biochem 168, 357–364.
Hanson, D.M., et al., 1969 J. Biol. Chem. 244, 2440–2449.
Holloman, W.K., et al., 1981 J. Biol. Chem. 256, 5087–5094.
Badman, R., 1972 Genetic Res., Camb. 20, 213–229.
Shenk, T.E., et al., 1975 Proc. Nat. Acad. Sci. USA, 72, 989–993.
Maekawa, K., et al., 1991 Eur. J. Biochem 200, 651–661.
Kowalski, D., et al, 1976 Biochemistry 15, 4457–4462.
Kroeker, W.D., et al., 1976 Biochemistry 15, 4463–4467.
Ardelt, W., et al., 1971 Biochem. Biophys. Res. Commun. 44, 1205–1211.
Kowalski, D., 1984 Nucleic Acids. Res. 12, 7071–7086.
Strickland, J.A., et al., 1991 Biochemistry 30, 9749–9756.
Doetsch, P.W., et al., 1988 Nucleic Acids Res. 16, 6935–6952.
Caron, P.R., et al., 1985 Proc. Nat. Acad. Sci. USA 82, 4925–4929.
Yeh, Y.–C., et al., 1994 J. Biol. Chem. 269, 15498–15504.
Welch, W.J., et al., 1983 J. Biol. Chem. 258, 7102–7111.
Jackson, S.P., et al., 1989 Proc. Nat. Acad. Sci. USA 86, 1781–1785.
Jackson, S.P., et al., 1988 Cell 55, 125–133.
He, J., et al., 1995 Nature 373, 721–724.
Peeples, M.E. 1988 Virology 162, 255–259.
Buckley, A., et al., 1988 J. Gen. Virology, 69, 1913–1920.
Dodgson et al. (1977) Biochemisty 16:2374–79.
Youil et al. (1996) Genomics 32:431–35.
Gauffre, A., et al., 1992 Molecular Immunology 29, 1113–1120.
Jones, B.K., et al., 1988 Proc. Natl. Acad. Sci. USA 85, 8410–8414.
Sarker, et al., 1992 Nucleic Acids Research 20: 871–878.
Meyers, R.M. et al., 1986 CSHSQB 51: 275.

Primary Examiner—W. Gary Jones
Assistant Examiner—Amy Atzel
Attorney, Agent, or Firm—Dann, Dorfman, Herrell and Skillman

[57] **ABSTRACT**

An endonuclease and its method of use for the detection of mutations in targeted polynucleotide sequences are provided, which facilitate the localization and identification of mutations, mismatches and genetic polymorphisms.

17 Claims, 13 Drawing Sheets



MISMATCH ENDONUCLEASE AND ITS USE IN IDENTIFYING MUTATIONS IN TARGETED POLYNUCLEOTIDE STRANDS

Pursuant to 35 U.S.C. §202(c), it is hereby acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Institute of Health.

FIELD OF THE INVENTION

This invention relates to materials and methods for the detection of mutations in targeted nucleic acids. More specifically, the invention provides a novel mismatch specific nuclease and methods of use of the enzyme that facilitate the genetic screening of hereditary diseases and cancer. The method is also useful for the detection of genetic polymorphisms.

BACKGROUND OF THE INVENTION

Several publications are referenced in this application by numerals in parenthesis in order to more fully describe the state of the art to which this invention pertains. Full citations for these references are found at the end of the specification. The disclosure of each of these publications is incorporated by reference in the present specification.

The sequence of nucleotides within a gene can be mutationally altered or "mismatched" in any of several ways, the most frequent of which being base-pair substitutions, frameshift mutations and deletions or insertions. These mutations can be induced by environmental factors, such as radiation and mutagenic chemicals; errors are also occasionally committed by DNA polymerases during replication. Many human disease states arise because fidelity of DNA replication is not maintained. Cystic fibrosis, sickle cell anemia and some cancers are caused by single base changes in the DNA resulting in the synthesis of aberrant or non-functional proteins.

The high growth rate of plants and the abundance of DNA intercalators in plants suggests an enhanced propensity for mismatch and frameshift lesions. Plants and fungi are known to possess an abundance of single-stranded specific nucleases that attack both DNA and RNA (9–14). Some of these, like the Nuclease α of *Ustilago maydis,* are suggested to take part in gene conversion during DNA recombination (15, 16). Of these nucleases, S1 nuclease from *Aspergillus oryzue* (17), and P1 nuclease from *Penicillium citrinum* (18), and Mung Bean Nuclease from the sprouts of *Vigna radiata* (19–22) are the best characterized. S1, P1 and the Mung Bean Nuclease are Zn proteins active mainly near pH 5.0 while Nuclease α is active at pH 8.0. The single strandedness property of DNA lesions appears to have been used by a plant enzyme, SP nuclease, for bulky adduct repair. The nuclease SP, purified from spinach, is a single-stranded DNase, an RNase, and able to incise DNA at TC_{6-4} dimers and cisplatin lesions, all at neutral pH (23, 24). It is not yet known whether SP can incise DNA at mismatches.

In *Escherichia coli,* lesions of base-substitution and unpaired DNA loops are repaired by a methylation-directed long patch repair system. The proteins in this multienzyme system include MutH, MutL and MutS (1, 2). This system is efficient, but the C/C lesion and DNA loops larger than 4 nucleotides are not repaired. The MutS and MutL proteins are conserved from bacteria to humans, and appear to be able to perform similar repair roles in higher organisms. For some of the lesions not well repaired by the MutS/MutL system, and for gene conversion where short-patch repair systems may be more desirable, other mismatch repair systems with novel capabilities are needed.

Currently, the most direct method for mutational analysis is DNA sequencing, however it is also the most labor intensive and expensive. It is usually not practical to sequence all potentially relevant regions of every experimental sample. Instead some type of preliminary screening method is commonly used to identify and target for sequencing only those samples that contain mutations. Single stranded conformational polymorphism (SSCP) is a widely used screening method based on mobility differences between single-stranded wild type and mutant sequences on native polyacrylamide gels. Other methods are based on mobility differences in wild type/mutant heteroduplexes (compared to control homoduplexes) on native gels (heteroduplex analysis) or denaturing gels (denaturing gradient gel electrophoresis). While sample preparation is relatively easy in these assays, very exacting conditions for electrophoresis are required to generate the often subtle mobility differences that form the basis for identifying the targets that contain mutations. Another critical parameter is the size of the target region being screened. In general, SSCP is used to screen target regions no longer than about 200–300 bases. The reliability of SSCP for detecting single-base mutations is somewhat uncertain but is probably in the 70–90% range for targets less than 200 bases. As the size of the target region increases, the detection rate declines, for example in one study from 87% for 183 bp targets to 57% for targets 307 bp in length (35). The ability to screen longer regions in a single step would enhance the utility of any mutation screening method.

Another type of screening technique currently in use is based on cleavage of unpaired bases in heteroduplexes formed between wild type probes hybridized to experimental targets containing point mutations. The cleavage products are also analyzed by gel electrophoresis, as subfragments generated by cleavage of the probe at a mismatch generally differ significantly in size from full length, uncleaved probe and are easily detected with a standard gel system. Mismatch cleavage has been effected either chemically (osmium tetroxide, hydroxylamine) or with a less toxic, enzymatic alternative, using RNase A. The RNase A cleavage assay has also been used, although much less frequently, to screen for mutations in endogenous mRNA targets for detecting mutations in DNA targets amplified by PCR. A mutation detection rate of over 50% was reported for the original RNase screening method (36).

A newer method to detect mutations in DNA relies on DNA ligase which covalently joins two adjacent oligonucleotides which are hybridized on a complementary target nucleic acid. The mismatch must occur at the site of ligation. As with other methods that rely on oligonucleotides, salt concentration and temperature at hybridization are crucial. Another consideration is the amount of enzyme added relative to the DNA concentration.

The methods mentioned above cannot reliably detect a base change in a nucleic acid which is contaminated with more than 80% of a background nucleic acid, such as normal or wild type sequences. Contamination problems are significant in cancer detection wherein a malignant cell, in circulation for example, is present in extremely low amounts. The methods now in use lack adequate sensitivity to be practically applied in the clinical setting.

A method for the detection of gene mutations with mismatch repair enzymes has been described by Lu-Chang and Hsu. See WO 93/20233. The product of the MutY gene

which recognizes mispaired A/G residues is employed in conjunction with another enzyme described in the reference as an "all type enzyme" which can nick at all base pair mismatches. The enzyme does not detect insertions and deletions. Also, the all type enzyme recognizes different mismatches with differing efficiencies and its activity can be adversely affected by flanking DNA sequences. This method therefore relies on a cocktail of mismatch repair enzymes and DNA glycosylases to detect the variety of mutations that can occur in a given DNA molecule.

Often, in the clinical setting, the nature of the mutation or mismatch is unknown so that the use of specific DNA glycosylases is precluded. Thus, there is a need for a single enzyme system that is capable of recognizing all mismatches with equal efficiency and also detecting insertions and deletions, regardless of the flanking DNA sequences. It would be beneficial to have a sensitive and accurate assay for detecting single base pair mismatches which does not require a large amount of sample, does not require the use of toxic chemicals, is neither labor intensive nor expensive and is capable of detecting not only mismatches but deletions and insertions of DNA as well.

Such a system, coupled with a method that would facilitate the identification of the location of the mutation in a given DNA molecule would be clearly advantageous for genetic screening applications. It is the purpose of the present invention to provide this novel mutation detection system.

SUMMARY OF THE INVENTION

The present invention provides materials and methods for the detection of mutations or mismatches in a targeted polynucleotide strand. Detection is achieved using a novel endonuclease in combination with a gel assay system that facilitates the screening and identification of altered base pairing in targeted nucleic acid strands.

According to one aspect of the invention, there is provided a novel nuclease, derived from celery and suitable for use in the detection of mutations or mismatches in target DNA or RNA. Celery (*Apium graveolens* var. dulce) contains abundant amounts of the nuclease of the invention which is highly specific for insertional/deletional DNA loop lesions and mismatches. This enzyme, designated herein as CEL I, incises at the phosphodiester bond at the 3' side of the mismatched nucleotide. CEL I has been purified about 10,000 fold, so as to be substantially homogeneous.

In a preferred embodiment of the invention, a method is provided for determining a mutation in a target sequence of single stranded mammalian polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including the target sequence. The sequences are amplified by polymerase chain reaction (PCR), labeled with a detectable marker, hybridized to one another, exposed to CEL I of the present invention, and analyzed on gels for the presence of the mutation.

The plant based endonuclease of the invention has a unique combination of properties. These include the ability to detect all possible mismatches between the hybridized sequences formed in performing the method of the invention; recognize polynucleotide loops and insertions between such hybridized sequences; detect polymorphisms between such hybridized strands; recognize sequence differences in polynucleotide strands between about 100 bp and 3 kb in length and recognize such mutations in a target polynucleotide sequence without substantial adverse effects of flanking DNA sequences.

The plant-based endonuclease, CEL I of the invention is not unique to celery. Similar enzymatic activities have been demonstrated in fourteen different plant species. Therefore, the enzyme is likely to be conserved in the plant kingdom and may be purified from plants other than celery. The procedure to purify this endonuclease activity from a plant other than celery is well known to those skilled in the art and enzymatic activity so isolated is contemplated to be within the scope of the present invention.

The plant-based endonuclease may not be limited to the plant kingdom but may be found in other life forms as well. Such enzymes may serve functions similar to that of CEL I in celery or be adapted for other special steps of DNA metabolism. Such enzymes or the genes encoding them may be used or modified to produce enzymatic activities that can function like CEL I. The isolation of such genes and their modification is also within the scope of the present invention.

In another embodiment of the invention, the above-described method is employed in conjunction with the addition of DNA ligase, DNA polymerase or a combination thereof thereby reducing non-specific DNA cleavage.

In yet another embodiment of the invention, the simultaneous analysis of multiple samples is performed using the above-described enzyme and method of the invention by a technique referred to herein as multiplex analysis.

In order to more clearly set forth the parameters of the present invention, the following definitions are used:

The term "endonuclease" refers to an enzyme that can cleave DNA internally.

The term "isolated nucleic acid" refers to a DNA or RNA molecule that is separated from sequences with which it is normally immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism in which it originates.

The term "base pair mismatch" indicates a base pair combination that generally does not form in nucleic acids according to Watson and Crick base pairing rules. For example, when dealing with the bases commonly found in DNA, namely adenine, guanine, cytosine and thymidine, base pair mismatches are those base combinations other than the A-T and G-C pairs normally found in DNA. As described herein, a mismatch may be indicated, for example as C/C meaning that a cytosine residue is found opposite another cytosine, as opposed to the proper pairing partner, guanine.

The phrase "DNA insertion or deletion" refers to the presence or absence of "matched" bases between two strands of DNA such that complementarity is not maintained over the region of inserted or deleted bases.

The term "complementary" refers to two DNA strands that exhibit substantial normal base pairing characteristics. Complementary DNA may contain one or more mismatches, however.

The term "hybridization" refers to the hydrogen bonding that occurs between two complementary DNA strands.

The phrase "flanking nucleic acid sequences" refers to those contiguous nucleic acid sequences that are 5' and 3' to the endonuclease cleavage site.

The term "multiplex analysis" refers to the simultaneous assay of pooled DNA samples according to the above described methods.

The term "substantially pure" refers to a preparation comprising at least 50–60% by weight of the material of interest. More preferably, the preparation comprises at least 75% by weight, and most preferably 90–99% by weight of

the material of interest. Purity is measured by methods appropriate for the material being purified, which in the case of protein includes chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis and the like.

C>T indicates the substitution of a cytosine residue for a thymidine residue giving rise to a mismatch. Inappropriate substitution of any base for another giving rise to a mismatch or a polymorphism may be indicated this way.

N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA) is a fluorescent dye used to label DNA molecular weight standards which are in turn utilized as an internal standard for DNA analyzed by automated DNA sequencing.

Primers may be labeled fluorescently with 6-carboxyfluorescein (6-FAM). Alternatively primers may be labeled with 4,7,2',7'-Tetrachloro-6-carboxyfluorescein (TET). Other alternative DNA labeling methods are known in the art and are contemplated to be within the scope of the invention.

CEL I has been purified so as to be substantially homogeneous, thus, peptide sequencing of the amino terminus is envisioned to provide the corresponding specific oligonucleotide probes to facilitate cloning of the enzyme from celery. Following cloning and sequencing of the gene, it may be expressed in any number of recombinant DNA systems. This procedure is well known to those skilled in the art and is contemplated to be within the scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** shows the results of sodium dodecyl sulfate (SDS) polyacrylamide gel analysis of the purified enzyme, CEL I. The positions of molecular weight markers are shown on the side. T indicates the top of the resolving gel.

FIG. **2** depicts certain heteroduplex DNA substrates used in performing nucleic acid analyses in accordance with the present invention. FIG. **2**A depicts a 64-mer which can be terminally labeled at either the 5'-P or the 3'-OH. The nucleotide positions used as a reference in this analysis are indicated irrespective of the number of nucleotide insertions at X in the top strand. The inserted sequences and substrate numbers are indicated in the table. FIG. **2**B illustrates mismatched basepair substrates used in this analysis, with the identities of nucleotides Y and Z varied as in the accompanying table to produce various mispaired substrates.

FIG. **3** is an autoradiogram demonstrating the effect of temperature on CEL I incisions in different substrates.

FIG. **4** is an autoradiogram illustrating the relative incision preferences of CEL I at DNA loops of one nucleotide. FIG. **4**A shows that in addition to the X=G, the X=C also allows two alternate basepairing conformations. FIG. **4**B demonstrates that the bottom strand of the substrate is competent for CEL I incision as in the C/C mismatch, #10, in lane **16**.

FIG. **5** is an autoradiogram of denaturing 15% polyacrylamide gels showing the AmpliTaq DNA polymerase mediated stimulation of purified CEL I incision at DNA mismatches of a single extrahelical nucleotide. F indicates the full length substrate, 64 nucleotide long, labeled at the 5' terminus (*) of the top strand. In panels **5**A, **5**B and **5**C, substrates were treated with varying quantities of CEL I in the presence or absence of DNA polymerase.

FIG. **6** is an autoradiogram showing the pH optimum of CEL I incision at the extrahelical G residue in the presence or absence of AmpliTaq DNA polymerase. The top panel shows the CEL I activity in the absence of AmpliTaq DNA polymerase. The bottom panel shows CEL I activity in the presence of polymerase.

FIG. **7** is an autoradiogram demonstrating the recognition of base substitution mismatches by purified CEL I in the presence of AmpliTaq DNA polymerase. (I) indicates the primary incision site at the phosphodiester bond 3' of a mismatched nucleotide. Panel **7**A illustrates cleavage of the substrate in the presence of both CEL I and DNA polymerase. In panel **7**B, CEL I was omitted.

FIG. **8** is an autoradiogram illustrating the ability of CEL I to recognize mutations in pooled DNA samples in the presence of excess wild-type DNA. Lanes **3, 5, 6, 10, 11, 12,** and **13** contain single samples containing wild type heteroduplexes. Lanes **4** and **6** contain an AG deletion. Lanes **8** and **9** contain a substrate with an 11 base-pair loop. The samples described above were pooled and treated with CEL I. The results of this "multiplex analysis" are shown in Lane **14**.

FIG. **9** is an autoradiogram further illustrating the ability of CEL I to recognize mutations in the presence of excess wild-type DNA. 1, 2, 3, 4, 10 or 30 heteroduplexed, radiolabeled PCR products (amplified from exon 2 of the BRCA1 gene) were exposed to CEL I in a single reaction tube and the products run on a 6% polyacrylamide gel. Lanes 1 and **2** are negative controls run in the absence of CEL I. Lane **3** to **11** contain 1 sample with the AG deletion in the presence of increasing amounts of wild-type non-mutated heteroduplexes.

FIG. **10** shows a schematic representative diagram of the BRCA1 gene and the exon boundaries in the gene. The sequence of BRCA1 is set forth as Sequence I.D. No. 1.

FIG. **11** is a histogram of a sample showing the localization of a 5 base deletion in the 11D exon of BRCA1 following PCR amplification and treatment with CEL I. A spike indicates a DNA fragment of a specific size generated by cleavage by CEL I at the site of a mismatch. Panel A shows the results obtained with a 6-FAM labeled primer annealed at nucleotide 3177 of BRCA1. Panel B shows the results obtained with a TET labeled primer annealed 73 bases into the intron between exon 11 and exon 12. Panel C represents the TAMRA internal lane size standard. Note that the position of the mutation can be assessed on both strands of DNA.

FIG. **12** is a histogram of a sample showing the localization of nonsense mutation, A>T, at position 2154 and a polymorphism C>T at nucleotide 2201 in the 11C exon of BRCA1 following PCR amplification and treatment with CEL I. Panel A shows a spike at base #700 and Panel B shows a spike at #305 corresponding to the site of the nonsense mutation. Panel C is the TAMRA internal lane standard.

FIG. **13** shows the results obtained from four different samples analyzed for the presence of mutations in exon 11A using the methods of the instant invention. Results from the 6-FAM samples are shown. Panel A shows a polymorphism T>C at nucleotide 2430 and a second spike at position #483 corresponding to the site of another polymorphism C>T at nucleotide 2731. Panel B shows only the second polymorphism described in panel A. Panel C shows no polymorphism or mutation. Panel D shows the two polymorphisms seen in panel A.

DETAILED DESCRIPTION OF THE INVENTION

The enzymatic basis for the maintenance of correct base sequences during DNA replication has been extensively

studied in *E. coli.* This organism has evolved a mismatch repair pathway that corrects a variety of DNA basepair mismatches in hemimethylated DNA as well as insertions/deletions up to four nucleotides long. Cells deficient in this pathway mutate more frequently, hence the genes are called MutS, MutL and MutH etc. MutS protein binds to the mismatch and MutH is the endonuclease that incises the DNA at a GATC site on the strand in which the A residue is not methylated. MutL forms a complex with MutH and MutS during repair. Homologs of MutS and MutL, but not MutH exist in many systems. In yeast MSH2 (MutS homolog) can bind to a mismatch by itself, but a complex of two MutL homologs (MLH and PMS1) plus a MSH2 has been observed. The human homolog hMSH2 has evolved to bind to larger DNA insertions up to 14 nucleotides in length, which frequently arise by mechanisms such as misalignment at the microsatelite repeats in humans. A role for hMLH1 in loop repair is unclear. Mutations in any one of these human homologs were shown to be responsible for the hereditary form of non-polyposis colon cancer (27, 28).

Celery contains over 40 μg of psoralen, a photoreactive intercalator, per gram of tissue (3). As a necessity, celery may possess a high capability for the repair of lesions of insertion, deletion, and other psoralen photoadducts. Single-strandedness at the site of the lesion is common to base substitution and DNA loop lesions. The data in the following examples demonstrate that celery possesses ample mismatch-specific endonuclease to deal with these potentially mutagenic events.

It has been found that the incision at a mismatch site by CEL I is greatly stimulated by the presence of a DNA polymerase. For a DNA loop containing a single nucleotide insertion, CEL I substrate preference is A≧G>T>C. For base-substitution mismatched basepairs, CEL I preference is C/C≧C/A~C/T≧G/G>A/C~A/A~T/C>T/G~G/T~G/A~A/G>T/T. CEL I shows a broad pH optimum from pH 6 to pH 9. To a lesser extent compared with loop incisions, CEL I is also a single-stranded DNase, and a weak exonuclease. CEL I possesses novel biochemical activities when compared to other nucleases. Mung Bean Nuclease is a 39 kd nuclease that is a single-stranded DNase and RNase, and has the ability to nick DNA at destabilized regions and DNA loops (19–22). However, it has a pH optimum at 5.0. It is not known whether Mung Bean Nuclease activity can be stimulated by a DNA polymerase as in the case of CEL I. Thus CEL I and Mung Bean Nuclease appear to be different enzymes; however this has not yet been conclusively confirmed.

The mechanism responsible for the AmpliTaq DNA polymerase stimulation of the CEL I activity is presently unknown. One possibility is that the DNA polymerase has a high affinity for the 3'—OH group produced by the CEL I incision at the mismatch and displaces CEL I simply by competition for the site. CEL I may have different affinities for the 3'—OH termini generated by incisions at different mismatches, thereby attenuating the extent that AmpliTaq DNA polymerase can stimulate its activity. The use of a DNA polymerase to displace a repair endonuclease in DNA repair was also observed for the UvrABC endonuclease mechanism (25). It was shown that the UvrABC endonuclease does not turnover unless it is in the presence of DNA polymerase I. The protein factors in vivo that can stimulate the CEL I activity may not be limited to DNA polymerases. It is possible that DNA helicases, DNA ligases, 3'-5' exonucleases or proteins that bind to DNA termini may perform that function.

It is important to note that a 5'-labeled substrate can be used to show a CEL I incision band in a denaturing polyacrylamide gel. Recently, a putative human all-type mismatch incision activity (24) was shown to be related to the human topoisomerase I. This enzyme is unable to release itself from a 5'-labeled substrate after mismatch nicking due to the formation of a covalent enzyme-DNA intermediate with the 3' terminus of the DNA nick (26). This covalent protein-DNA complex cannot migrate into the denaturing polyacrylamide gel to form a band. CEL I mismatch nicking has been demonstrated with 5' labeled substrates. Therefore, CEL I is not a plant equivalent of the topoisomerase I-like human all-type mismatch repair activity.

CEL I appears to be a mannopyranosyl glycoprotein as judged by its tight binding to Concanavalin A-Sepharose resin and by the staining of CEL I with the Periodic acid-Schiff glycoprotein stain. Insofar as it is known, no repair enzyme has been demonstrated to be a glycoprotein. Glycoproteins are often found to be excreted from the cell, on cellular membranes or secreted into organelles. However, glycoproteins have also been shown to exist in the nucleus for important functions. The level of a 100 kDa stress glycoprotein was found to increase in the nucleus when Gerbil fibroma cells are subjected to heat shock treatment (27). Transcription factors for RNA polymerase II in human cells are known to be modified with N-acetylglucosamine residues (28, 29). Recently, lactoferrin, an iron-binding glycoprotein, was found to bind to DNA in the nucleus of human cells and it activated transcription in a sequence-specific manner (30). The nuclei of cells infected with some viruses are known to contain viral glycoproteins (31–33). These examples where glycoproteins are known to exist inside the nucleus, not merely on the nuclear membrane or at the nuclear pores, tend to show that glycosylated proteins may be important in the nucleus. CEL I appears to be an example of a glycoprotein that can participate in DNA repair.

The properties of the celery mismatch endonuclease CEL I resemble those of single-stranded nucleases. The best-suited substrates for CEL I are DNA loops and base-substitution mismatches such as the C/C mismatch. In contrast, loops greater than 4 nucleotides and the C/C mismatch are the substrates worst-suited for the *E. coli* mutHLS mismatch repair system (1, 2). Thus CEL I is an enzyme that possesses novel mismatch endonuclease activity.

The following examples are provided to describe the invention in further detail. These examples, which set forth the best mode presently contemplated for carrying out the invention, are intended to illustrate and not to limit the invention.

EXAMPLE I

Purification of CEL I

Two different CEL I preparations were made up as described below. Their properties are similar except that the less pure preparation (Mono Q fraction) may contain protein factors that can stimulate the CEL I activity.

(i) Preparation of CEL I Mono Q fraction

100 gm of celery stalk was homogenized in a Waring blender with 100 ml of a buffer of 0.1M Tris-HCl pH 7.0 with 10 μM phenylmethanesulfonyl fluoride (PMSF) (Buffer A) at 4° C. for 2 minutes. The mixture was cleared by centrifugation, and the supernatant was stored at −70° C. The extract was fractionated by anion exchange chromatography on a FPLC Mono Q HR5/10 column. The bound CEL I nuclease activity was eluted with a linear gradient of salt at about 0.15M KCl.

(ii) Preparation of highly purified CEL I

7 Kg of celery at 4° C. was extracted with a juicer and adjusted with 10X Buffer A to give a final concentration of 1X Buffer A. The extract was concentrated with a 25% to 85% saturation ammonium sulfate precipitation step. The final pellet was dissolved in 250 ml of Buffer A and dialyzed against 0.5M KCl in Buffer A. The solution was incubated with 10 ml of Concanavalin A-Sepharose resin (Sigma) overnight at 4° C. The slurry was packed into a 2.5 cm diameter column and washed with 0.5M KCl in Buffer A. The bound CEL I was eluted with 60 ml of 0.3M α-D mannose, 0.5M KCl in Buffer A at 65° C. The CEL I was dialyzed against a solution of 25 mM KPO_4, 10 μM PMSF, pH 7.4 (Buffer B), and applied to a phosphocellulose column that had been equilibrated in the Buffer B. The bound enzyme was eluted with a linear gradient of KCl in Buffer B. The peak of CEL I activity from this column was further fractionated by size on a Superose 12 FPLC column in 0.2M KCl, 1 mM $ZnCl_2$, 10 μM PMSF, 50 mM Tris-HCl pH 7.8. The center of the CEL I peak from this gel filtration step was used as the purified CEL I in this study. A protein band of about 34,000 daltons is visible when 5 micrograms of CEL I of the Superose 12 fraction was visualized with Coomassie Blue staining or carbohydrate staining (Periodic acid-Schiff base mediated staining kit, SIGMA Chemicals (5)) on a 15% polyacrylamide SDS PAGE gel as shown in FIG. **1**. A second band of approximately 36,000 daltons was also visible in the gel. Both bands were stained with the glycoprotein specific stain. The subtle mobility differences observed in the two bands may be due to differential glycosylation. Alternatively, there may be a contaminant in the preparation which co-purifies with CEL I.

Protein determination

Protein concentrations of the samples were determined by the Bicinchoninic acid protein assay (4, Pierce).

Following purification of CEL I enzyme, mutational analysis on experimental and clinical DNA substrates were performed in a suitable gel system. CEL I recognized and cleaved DNA at a variety of mismatches, deletions and insertions. The following examples describe in greater detail the manner in which mutational analysis is practiced according to this invention.

EXAMPLE II

Preparation of heteroduplexes containing various mismatches

DNA heteroduplex substrates of 64 basepairs long were constructed containing mismatched basepairs or DNA loops which were prepared using similar methods reported in Jones and Yeung (34). The DNA loops are composed of different nucleotides and various loop sizes as illustrated in FIG. **2**. The DNA duplexes were labeled at one of the four termini so that DNA endonuclease incisions at the mispaired nucleotides could be identified as a truncated DNA band on a denaturing DNA sequencing gel. The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and purified by using a denaturing PAGE gel in the presence of 7M urea at 50° C. The purified single-stranded oligonucleotides were hybridized with appropriate opposite strands. The DNA duplex, containing mismatches or not, was purified by using a nondenaturing PAGE gel. DNA was eluted from the gel slice by using electro-elution in a Centricon unit in an AMICON model 57005 electroeluter. The upper reservoir of this unit has been redesigned to include water-tight partitions that prevent cross-contamination. The sequences of the substrates used are set forth below:

SEQ. I.D. No. 2 is the top strand of Substrate Nos. 1, 12, 13, and 14: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGCGTCAT GTGTGGAATT GTGATTAAAA TCG-3';

SEQ. I.D. No. 3 is the bottom strand of Substrate Nos. 1, 2, 3, 4, 5, 7, 10, 15:

5'-GCGATTTTAA TCACAATTCC ACACATGACG CGAGCCGGAA GCATAAAGTG, AACTAGCATG ACG-3';

SEQ. I.D. No. 4 is the top strand of Substrate No. 2: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGGCGTCA TGTGTGGAAT TGTGATTAAA ATCG-3';

SEQ. I.D. No. 5 is the top strand of Substrate No. 3: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGTCGTCA TGTGTGGAAT TGTGATTAAA ATCG-3';

SEQ. I.D. No. 6 is the top strand of Substrate No. 4: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGACGTCA TGTGTGGAAT TGTGATTAAA ATCG-3';

SEQ. I.D. No. 7 is the top strand of Substrate No. 5: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGCCGTCA TGTGTGGAAT TGTGATTAAA ATCG-3';

SEQ. I.D. No. 8 is the top strand of Substrate Nos. 6, 7, 8, 18: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCACGTCAT GTGTGGAATT GTGATTAAAA TCG-3';

SEQ. I.D. No. 9 is the top strand of Substrate Nos. 9, 10, 11, 19: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCCCGTCAT GTGTGGAATT GTGATTAAAA TCG-3';

SEQ. I.D. No. 10 is the top strand of Substrate Nos. 15, 16, 17, 20: 5'-CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCTCGTCAT GTGTGGAATT GTGATTAAAA TCG-3';

SEQ. I.D. No. 11 is the bottom strand of Substrate Nos. 6, 9, 12, 20: 5'-GCGATTTTAA TCACAATTCC ACACATCACG AGAGCCGGAA GCATAAAGTG AACTAGCATG ACG-3';

SEQ. I.D. No. 12 is the bottom strand of Substrate Nos. 8, 13, 16, 19: 5'-GCGATTTTAA TCACAATTCC ACACATCACG GGAGCCGGAA GCATAAAGTG AACTAGCATG ACG-3';

SEQ. I.D. No. 13 is the bottom strand of Substrate Nos. 11, 14, 17, 18: 5'-GCGATTTTAA TCACAATTCC ACACATCACG TGAGCCGGAA GCATAAAGTG AACTAGCATG ACG-3'.

EXAMPLE III

Mismatch endonuclease assay

Fifty to 100 fmol of 5 ' [^{32}P]-labeled substrate described in Example II were incubated with the Mono Q CEL I preparation in 20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$ for 30 minutes at temperatures of 0° C. to 80° C. From one half to 2.5 units of AmpliTaq DNA polymerase was added to the nuclease assay reaction. Ten μM dNTP was included in the reaction mixture where indicated (FIGS. **2** & **5**). The 20 μL reaction was terminated by adding 10 μL of 1.5% SDS, 47 mM EDTA, and 75% formamide plus tracking dyes and analyzed on a denaturing 15% PAGE gel in 7M urea at 50° C. An autoradiogram was used to visualize the radioactive bands. Chemical DNA sequencing ladders were included as size markers. Incision sites were accurately determined by co-electrophoresis of the incision band and the DNA sequencing ladder in the same lane.

EXAMPLE IV

The Effect of Temperature on CEL I Incision Activity at single-nucleotide DNA loop and nucleotide substitutions

The CEL I fraction eluted from the Mono Q chromatography of the celery extract was found to specifically nick DNA heteroduplexes containing DNA loops with a single extrahelical guanine (substrate #2) or thymine residue (#3), but not the perfectly basepaired DNA duplex #1 as shown in FIG. **3**. In these experiments fifty fmol of heteroduplex #2 (lanes **3–9**), #3 (lanes **10–16**), perfectly basepaired duplex #1 (lanes **17–23**) and single-stranded DNA substrate (lanes **24–30**), each labeled at the 5'-terminus with γ-[^{32}p] ATP and T4 polynucleotide kinase at about 6000 Ci/mmol, were incubated with 0.5 μL (10 μg) of the Mono Q fraction of the CEL I preparation in 20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$ for 30 minutes at various temperatures. Each 20 μL reaction was terminated by adding 10 μL of 1.5% SDS, 47 mM EDTA, and 75% formamide containing xylene cyanol and bromophenol blue. Ten μL of the sample was loaded onto a 15% polyacrylamide, 7M urea denaturing DNA sequencing gel at about 50° C., and subjected to electrophoretic separation and autoradiography as previously reported (7). The G+A and the T chemical sequencing reactions were performed as described (7) and used as size markers. CEL I incision produced bands at about 35 nucleotides long. Lines are drawn from the positions of the incision bands to the phosphodiester bonds (I and II) nicked by the endonuclease in the reference sequencing ladder. For a 5'-labeled substrate, when a nuclease nicks 5' of a nucleotide and produces a 3'-OH terminus, the truncated band runs half a nucleotide spacing slower than the band for that nucleotide in the chemical DNA sequencing reaction product lane (**34**).

Substrate #2 can basepair in two conformations because the inserted G is within a CGCG sequence. Therefore either the G residue in the second or the third nucleotide position can become unpaired, possibly extrahelical in conformation, when this duplex is hybridized:

5'-CGGCG-3' or 5'-CGGCG-3'

3'-G-CGC-5' 5'-GC-GC-5'

Accordingly, two mismatch incision bands were observed, each correlating to the phosphodiester bond immediately 3' of the unpaired nucleotide. See FIG. **3**, lanes **3–9**. This slippage can occur in the target sequence only when G or C is in the mismatched top strand. Therefore, the non-paired T residue in substrate #3 gave one incision band at the same relative position as the upper band derived from the substrate #2. See FIG. **3**, lanes **10–16**. These gel mobilities are consistent with the production of a 3'-OH group on the deoxyribose moiety (6). CEL I increases in activity with temperature up to 45° C. as illustrated by the increase in band intensity, see FIG. **3**. However, from 65° C. to 80° C., specificity is diminished due to DNA duplex denaturation.

EXAMPLE V

Relative Incision Preferences of CEL I

To ascertain whether there is a single endonuclease incision at each DNA duplex, the experiment described in FIG. **3** was repeated with DNA labeled on the 3' terminus of the top strand. If there were only one incision site, initial incision positions revealed by substrates labeled at the 5' or the 3' termini should be at the same phosphodiester bond. In these experiments, substrates were labeled at the 3' termini with [^{32}P] α-dCTP, cold dGTP and the Klenow fragment of DNA polymerase I to about 6000 Ci/mmol. The sample preparation, denaturing gel resolution and autoradiogram analysis are the same as described in FIG. **3** except incubation of 50 fmole of substrate with 10 μg of the CEL I Mono Q fraction was for 30 minutes at a single temperature, 37° C. The DNA sequencing ladders for substrates #4 and #5 are shown in lanes **1–4** to illustrate the DNA sequences used. Lanes **5–8** had no enzyme during the incubation. Lanes **9–12** are mismatch endonuclease incisions of the substrates #2, #4, #5, #3, respectively. A line is drawn from the position of the incision band to the phosphodiester bond (I) nicked by the endonuclease in the reference sequencing ladder. Lanes **13** and **14** demonstrate the coelectrophoresis of the CEL I incision band with a chemical DNA sequencing ladder to accurately determine the incision position. Relative incision preferences for substrates #2, #3, #4, and #5 are shown in FIG. **4** for the 3' labeled substrates. The mobilities of the incision bands in lanes **9–12** of FIG. **4** indicate that the incision reactions had occurred at the phosphodiester bond immediately 3' of the unpaired nucleotide. Therefore, the incision site is the same for substrates labeled either at the 5' or the 3' terminus. The fact that the DNA incision was found to occur at the same bond position, whether the substrate DNA was labeled at the 5' termini or the 3' termini shows that CEL I is not a DNA glycosylase. A DNA glycosylase mechanism would cause the DNA incision position in the two DNA substrates to be one base apart because a base is excised by the DNA glycosylase.

Precise determination of the incision site was performed as in the example in lane **14** in which the T residue chemical sequencing reaction of the labeled top strand of substrate #2 (lane **13**) was mixed with the CEL I incision product of lane **9** and analyzed in the same lane. For a 3'-labeled substrate, when a nuclease nicks 3' of a nucleotide and produces a 5' PO_4 terminus, the truncated band runs with the band for that nucleotide in the chemical DNA sequencing reaction product lane (**7**). Moreover, the gel mobility, relative to the size standards of chemical DNA sequencing, illustrated that the DNA nick produced a 5'-phosphorylated terminus (6). For a DNA loop with a single nucleotide insertion, the nuclease specificity is A≧G>T>C. It can be seen in FIG. **4**A that a small amount of 5' to 3' exonuclease activity is present in this CEL I preparation.

To test whether CEL I can cut in the bottom strand across from a DNA loop of one nucleotide in the top strand, or whether nicking of the loop-containing strand may lead to secondary CEL I incision across from the nick, the bottom strand that contains no unpaired nucleotides in substrate #2 was labeled at the 3' end and incubated in the presence of CEL I. The extrahelical nucleotide in the top strand, or the DNA nick made by CEL I in the top strand of substrate #2, seen in lane **9** of FIG. **4**, did not lead to significant nicking of the bottom strand (lane **18**). As a control against the possibility that DNA sequence effect may favor CEL I incision in the top strand and not the bottom strand, CEL I was tested for incision of the bottom strand in the C/C mismatch substrate in lanes **15** and **16**. Mismatch incision was made when CEL I was present in lane **16**.

In the characterization of the incision site of a repair endonuclease, it is important to determine whether one or two incisions have been made for each lesion. This is normally accomplished by using lesion-containing substrates that have been labeled, in turn, at the four termini of a DNA duplex. This test has been satisfied in the analysis of substrate #2 by using three labeled substrates because of the near absence of incision in the bottom strand. In FIG. **3**, lane

4–7 and FIG. 4, lane 9, respectively, the incision of this substrate in both the 5' labeled and the 3' labeled substrates have been compared. The incision site was found to be at the 3' side of the mismatched nucleotide in both cases. The lack of incision on the bottom strand for substrate #2 was demonstrated in lanc 18 of FIG. 4. Only the 5' labeled substrate was needed in this case since no significant bottom strand incision had occurred.

EXAMPLE VI

Effect of AmpliTaq DNA polymerase on the incisions at DNA loop mismatches

CEL I activity is stimulated by the presence of a DNA polymerase. In FIG. 5, the CEL I incisions at single-nucleotide loop substrates were stimulated by AmpliTaq DNA polymerase to different extents depending on which nucleotides are present in the loop. It was necessary to use different amounts of CEL I to illustrate the AmpliTaq DNA polymerase stimulation. The stimulation of the incision at extrahelical C and extrahelical T substrates are best illustrated in FIGS. 5 A & B (compare lanes 4 with lanes 9, and lanes 5 with lanes 10, in the respective panels) where higher CEL I levels are required to show good incision at these mismatches. For extrahelical G and extrahelical A substrates that are among the best substrates for CEL I, AmpliTaq DNA polymerase stimulation can best be illustrated by using a much lower level of CEL I as in FIG. 5. The amounts of AmpliTaq stimulation of CEL I in FIG. 5 were quantified and presented in Table I.

TABLE I

Quantification of the CEL I incision bands shown in the autoradiogram in FIG. 5.

	–		+		
AmpliTaq Substrate	Panel lanc#	Counts	Panel lanc#	Counts	+/–
Extrahelical G,band I	A,2	20894	A,7	22101	1.1
Extrahelical A, band I	A,3	19451	A,8	26357	1.4
Extrahelical C, band I	A,4	4867	A,9	12009	2.5
Extrahelical T, band I	A,5	2297	A,10	25230	11.0
Extrahelical G, band I	B,2	12270	B,7	19510	1.6
Extrahelical A, band I	B,3	10936	B,8	24960	2.3
Extrahelical C, band I	B,4	1180	B,9	2597	2.2
Extrahelical T, band I	B,5	700	B,10	21086	30.1
Extrahelical G, band I	C,11	10409	C,13	18649	1.8
Extrahelical G, band II	C,11	9020	C,13	19912	2.2
Extrahelical A, band I	C,12	7165	C,14	14983	2.1

The Autoradiograms were quantified in two dimensions with an AMBIS densitometer and the amount of signal in each band is given as counts.

EXAMPLE VII

Optimum pH of CEL I Activity

The pH optimum of CEL I for the extrahelical G substrate was investigated in the absence or presence of the AmpliTaq DNA polymerase. CEL I (9.5 ng) was incubated with 100 fmol of the substrate in a 20 μL reaction in buffers of pH 5–6.5 (imidazole) and pH 7–9.5 (Tris-HCl) for 30 minutes at 37° C. When used, one half unit of AmpliTaq DNA polymerase was present in the incubation in the top (–polymerase) or bottom panels (+polymerase), respectively. As shown in FIG. 6, CEL I was found to be active from pII 5.0 to pH 9.5, and showed a broad pH optimum centered about pH 7.5 (top panel). When AmpliTaq DNA polymerase was present, the incision was stimulated across the whole pH range (bottom panel). The assay method did not use initial kinetics and thus precluded quantitative conclusions on this pH profile of CEL I. However, it is clear that the enzyme works very well in the neutral pH ranges.

EXAMPLE VIII

Incisions by CEL I at basepair substitutions

Other combinations of mismatched substrates are also recognized by CEL I and incised on one of the two DNA strands of each DNA duplex. Some of these substrates are less efficiently incised compared with those containing DNA loops; therefore 45° C. was used for incubation instead of 37° C. Substrates with the 5' termini of the top strands labeled were used in this study. The autoradiogram of FIG. 7 shows that mismatches containing a C residue are the preferred mismatch substrates with C/C often better than C/A and C/T. The incisions at these mismatches tend to produce two alternate incision positions, one at the phosphodiester bond 3' of the mismatched C residue, one at the phosphodiester bond one nucleotide further removed in the 3' direction. Whether alternate incision sites will be observed for these mismatches within another DNA sequence context has not been investigated. One possible explanation for this phenomenon may be greater basepair destabilization next to a mismatch that contains a C residue than for other base-substitutions. Alternatively, the specific mismatched nucleotide may shift one position to the 3' side because the next nucleotide is also a C residue and the two residues can exchange their roles in the pairing with the G residue in the opposite DNA strand. For base substitution mismatched basepairs, CEL I specificity in the presence of AmpliTaq DNA polymerase, with respect to the top strand, is C/C≧C/A~C/T≧G/G>A/C~A/A~T/C>T/G~G/T~G/A~A/G>T/T (FIG. 7A). Because eubacterial DNA polymerases are known to incise at unusual DNA structures (8), a test was conducted to determine whether AmpliTaq DNA polymerase by itself will incise at the 13 substrates used in FIG. 7. Under extended exposure of the autoradiogram, no mismatch incision by the AmpliTaq DNA polymerase was observed (FIG. 7B).

EXAMPLE IX

Detection of DNA mutations Using CEL-I and Multiplex Analysis

The sensitivity of CEL I for mismatch detection is illustrated by its ability to detect mutations in pooled DNA samples. DNA was obtained from peripheral blood lymphocytes from individuals undergoing genetic screening at the Fox Chase Cancer Center. Samples were obtained from breast cancer-only, ovarian cancer-only, breast/ovarian cancer syndrome families or from non-breast/ovarian cancer control samples. Unlabeled primers specific for exon 2 of BRCA1 were utilized to PCR amplify this region of the gene. The wild-type PCR products of exon 2 were labeled with gamma ^{32}P-ATP. Briefly, 10 picomoles of PCR product were purified by the Wizard procedure (Promega). Exon 2 wild-type products were then phosphorylated using T4 kinase and 15 picomoles of gamma ^{32}P-ATP at 6,000 Ci/mmol in 30 μl 1X kinase buffer (70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol) at 37° C. for 1 hour. The reactions were stopped with 1 μl 0.5M EDTA. The reaction volume was brought up to 50 μl with 1×STE buffer (100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM EDTA) and processed through a Pharmacia Probe Quant column. Labeled DNA (1 pmol/μl in 100 μl) was then used for

hybridization with individual unlabeled PCR amplified experimental samples. For each individual sample, 100 fmol of the unlabeled PCR amplified product was incubated with 200 fmol of the ^{32}P-labeled wild-type PCR product in CEL I reaction buffer (25 mM KCl, 10 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5). Following denaturation and renaturation, heteroduplexed, radiolabeled PCR products were exposed to CEL I for 30 minutes at 37° C. in 1X CEL reaction buffer and stopped via the addition of 10 μl stop mix (75% formamide, 47 mM EDTA, 1.5% SDS, xylene cyanol and bromophenol blue). The heteroduplexes were treated with the enzyme individually (lanes **4–13**) or pooled in one sample tube (lane **14**) and treated. The products of the reaction were loaded onto a 15% polyacrylamide gel containing 7M urea and the results are shown in FIG. **8**. Out of the 10 samples analyzed, 2 contained an AG deletion (lanes **4** and **7**), 2 contained an 11 base-pair loop (lanes **8** and **9**), and the other 6 were wild type (lanes **5, 6, 10, 11, 12,** and **13**). Cleavage by CEL I at the AG deletion resulted in the formation of two bands, one of approximately 151 nucleotides from the top strand, the other at 112 nucleotides from the bottom strand (lanes **4** and **7**). Cleavage by CEL I at 11 base-pair loops resulted in the formation of one band at 147 nucleotides from the top strand, and a group of bands at 109 nucleotides in the bottom strand (lanes **8** and **9**). Lanes **1, 2** and **3** contain DNA that was not exposed to CEL I as negative controls, lane **15** contains 64 and 34 bp nucleotide markers. As can be seen in lane **14** of the gel, when the samples were pooled and exposed simultaneously to CEL I, the enzyme cleaved at all of the above listed mutations with no loss of specificity. Also, the PCR products of the wild-type samples showed no non-specific DNA nicking.

To further illustrate the ability of CEL-I to detect mutations in pooled DNA samples, 1, 2, 3, 5, 10 or 30 heteroduplexed, radiolabelled PCR products, (again amplified from exon 2 of the BRCA1 gene), were exposed to CEL-I in a single reaction tube and the products run on a 6% polyacrylamide gel containing 7M urea. Samples were amplified and radiolabeled as described above. Each pool contained only one sample which had a mutation (AG deletion). The other samples in each pool were wild-type. Lanes **1** and **2** contain control samples which were not exposed to CEL I. In the pooled samples where a mutation was present, CEL-I consistently cleaved the PCR products illustrating the sensitivity of the enzyme in the presence of excess wild-type, non-mutated DNA (Lanes **4, 5, 6, 7, 8, 9,** and **11**). As a control, heteroduplexed PCR products containing no mutations were analyzed and no cut band corresponding to a mutation appeared (FIG. **9**, lanes **3** and **10**).

EXAMPLE X

Detection of Mutations and Polymorphisms by CEL-I in Samples Obtained from High Risk Families

PCR primer sets specific for the exons in the BRCA1 gene have been synthesized at Fox Chase Cancer Center. The gene sequence of BRCA1 is known. The exon boundaries and corresponding base numbers are shown in table II. Primers to amplify desired sequences can be readily designed by those skilled in the art following the methodology set forth in *Current Protocols in Molecular Biology,* Ausubel et al., eds, John Wiley and Sons, Inc. (1995). These primers were planned such than in each PCR reaction, one primer is labeled at the 5' termini with a fluorescent-label, 6-FAM, while the other primer is similarly labeled with a label of another color, TET. A PCR product will thus be labeled with two colors such that DNA nicking events in either strand can be observed independently and the measurements corroborated. A summary of the results is presented in Table III.

TABLE II

EXON BOUNDARIES AND CORRESPONDING BASE NUMBERS IN BRCA1

EXON	BASE #'s
1	1–100
2	101–199
3	200–253
5	254–331
6	332–420
7	421–560
8	561–665
9	666–712
10	713–788
11	789–4215
11B	789–1591
11C	1454–2459
11A	2248–3290
11D	3177–4215
12	4216–4302
13	4303–4476
14	4477–4603
15	4604–4794
16	4795–5105
17	5106–5193
18	5194–5273
19	5274–5310
20	5311–5396
21	5397–5451
22	5452–5526
23	5527–5586
24	5587–5711

FIG. **10** depicts a schematic of the exons present in the BRCA1 gene. Peripheral blood samples from individuals in high risk families were collected and the DNA isolated. The PCR products were amplified using Elongase (BRL) and purified using Wizard PCR Preps (Promega). The DNA was heated to 94° C. and slowly cooled in 1X CEL I buffer (20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$) to form heteroduplexes. The heteroduplexes were incubated in 20 μl 1X CEL I buffer with 0.2 μl of CEL I and 0.5 units of AmpliTaq at 45° C. for 30 minutes. The reactions were stopped with 1 mM phenanthroline and incubated for an additional 10 minutes at 45° C. The sample was processed through a Centricep column (Princeton Separations) and dried down. One microliter of ABI loading buffer (25 mM EDTA, pH 8.0, 50 mg/ml Blue dextran), 4 μl deionized formamide and 0.5 μl TAMRA internal lane standard were added to the dried DNA pellet. The sample was heated at 90° C. for 2 minutes and then quenched on ice prior to loading. The sample was then loaded onto a 4.25% denaturing 34 cm well-to-read acrylamide gel and analyzed on an ABI 373 Sequencer using GENESCAN 672 software. The 6-FAM labelled primer in this experimental sample was at nucleotide 3177 of the BRCA1 cDNA (region 11D), the TET labelled primer was 73 nucleotides into the intron between exon 11 and exon 12. Each spike represents the presence of a DNA band produced by the cleavage of the heteroduplex by CEL-I where a mutation or a polymorphism is present. One spike represents the size of the CEL I produced fragment from the 3' side of the mismatch site to the 5' 6-FAM

label of the top strand. The other spike represents the corresponding fragment in the bottom strand from the 3' side of the mismatch to the 5' TET label. The sum of the two fragments equals one base longer than the length of the PCR product. The 6-FAM panel shows a spike at base #645 from the 6-FAM label and the TET panel shows a spike at base #483 from the TET label, both corresponding to the site of the 5 base deletion at nucleotide 3819 of the BRCA1 cDNA (FIG. **11**).

Analysis of exon 11 in another individual was performed using a 6-FAM-labelled primer at nucleotide 1454 of the BRCA1 cDNA (FIG. **12**). The TET-labelled primer was at nucleotide 2459 (region 11C). The PCR amplified products were amplified and prepared as described above. In this individual, the 6-FAM panel shows a spike at base #700 and the TET panel shows a spike at #305, each spike corresponding to the site of CEL I incision in the respective DNA strand at a nonsense mutation of A>T at nucleotide 2154 of the BRCA1 cDNA. The 6-FAM panel also shows a spike at base #747 and the TET panel shows a spike at #258 corresponding to the site of a polymorphism C>T at nucleotide 2201 of the BRCA1 cDNA. The nonsense mutation and polymorphism have been confirmed by sequencing of this particular sample (KO-11) using the ABI 377 Sequencer. Spikes that are marked with an asterisk are also present in the no enzyme control lane and represent PCR product background.

Certain individuals have mutations in another region of exon 11, region 11A, on the schematic in FIG. **10**. A 6-FAM-labelled primer at nucleotide 2248 of the BRCA1 cDNA and a TET labeled primer at nucleotide 3290 were used to amplify this region of exon 11. Following amplification, the samples were processed as described above. The four 6-FAM panels represent CEL-I reactions with 4 different individual samples. The first panel in FIG. **13A**, sample #KO-2, shows one spike at #182 corresponding to the site of a polymorphism T>C at nucleotide 2430 and a second spike at nucleotide #483 corresponding to the site of another polymorphism C>T at nucleotide 2731. The second panel, FIG. **13B**, sample #KO-3, shows only the second polymorphism. The third panel, FIG. **13C**, sample #KO-7 shows no polymorphism. The fourth panel, FIG. **13D**, sample #KO-11, shows two spikes corresponding to the two polymorphisms. It is interesting to note that this sample, KO-11, shows up positive for a nonsense mutation and a polymorphism in the region of exon 11C corresponding to nucleotides 1454–2459 as described above.

TABLE III

SUMMARY OF BRCA1 MUTATIONS AND POLYMORPHISMS DETECTED BY CEL I

EXON	NUCLEOTIDE POSITION #	TYPE OF MUTATION
2	185	AG deletion
2	188	11 base deletion
11 C	2154	A > T
11 D	3819	5 base deletion
11 D	4168	A > G
11 D	4153	A deletion
11 D	4184	4 base deletion
20	5382	C insertion

EXON	NUCLEOTIDE POSITION #	TYPE OF POLYMORPHISM
11 B	1186	A > G
11 C	2201	T > C
11 A	2430	T > C

TABLE III-continued

SUMMARY OF BRCA1 MUTATIONS AND POLYMORPHISMS DETECTED BY CEL I

11 A	2731	C > T
11 D	3667	A > G

Table IV sets forth the 5' and 3' flanking sequences surrounding the mutations detected by CEL I in the present invention. While not exhaustive, it can be seen from the variability of the flanking sequences surrounding these mutations and polymorphisms that CEL I sensitivity and recognition of mismatched DNA heteroduplexes does not appear to be adversely affected by flanking sequences.

TABLE IV

EFFECT OF FLANKING SEQUENCES ON ENDONUCLEASE ACTIVITY OF CEL I

nucleotide position	EXON	type of change	5' flanking sequence	3' flanking sequence
185	2	AG deletion	5'ATCTT TAGGA3'	5' AGTGT TCACA 3'
188	2	11 bp deletion	5' TTAGA AATCT3'	5'G the next 4 bp are in intron
1186	11 B	A—> G	5' TAAGC ATTCG 3'	5' GAAAC CTTG 3'
2154	11 C	A—> T	5' GAGCC CTCGG 3'	5' AGAAG TCTTC 3'
2201	11 C	T--> C	5' GACAG CTGTC 3'	5' GATAC CTATG 3'
2430	11 A	T--> C	5' AGTAG TCATC 3'	5' AGTAT TCATA 3'
2731	11 A	C--> T	5' TGCTC ACGAG 3'	5' GTTTT CAAAA 3'
3667	11 D	A--> G	5' CAGAA CTCTT 3'	5' GGAGA CCTCT 3'
3819	11 D	5 bp deletion	5' GTAAA CATTT 3'	5' CAATA GTTAT 3'
4153	11 D	A deletion	5' TGATG ACTAC 3'	5' AGAAA TCTTT 3'
4184	11 D	4 bp deletion	5' AATAA TTATT 3'	5' GAAGA CTTCT 3'
4168	11 D	A--> G	5' AACGG TTGCC 3'	5' CTTGA GAACT 3'
5382	20	C insertion	5' ATCCC TAGGG 3'	5' AGGAC TCCTG 3'

As can be seen from the above described examples, utilization of CEL I has distinct advantages over methods employing other mismatch repair systems during analysis of mutations in the clinical setting. These advantages are summarized in Table V.

TABLE III

Comparison of the advantages of methods employing CEL I over current mismatch detection methods:

	S1 nuclease method (7)	DNA mismatch glycosylases (8)	MutS binding assay (9)	Chemical cleavage method (10)	T4 endo-nuclease VII (11)	RNase nicking mismatched RNA:DNA (12)	Automated DNA sequencing	ddNTP SSCP finger-printing	Plant mismatch endo-nuclease CEL I
Assay at neutral pH	no	yes	yes	yes	yes	yes	yes	yes	yes
Applicable to mutations of unknown positions	yes	no	yes	yes	yes	yes	yes	yes	yes
Applicable to all basepair substitutions	unknown	with difficulty	with difficulty	with difficulty	yes	no	yes	yes	yes
Applicable to DNA loops,	yes	no	with difficulty	multiple bands	yes	unknown	yes	yes	yes
Advantage of single major band in loop detection	no	no	yes	no	yes	no	no	no	yes
Advantage of little influence by sequence specificity	no	unknown	yes	unknown	cuts w/o mismatch	unknown	no	with difficulty	yes
Advantage of no RNA instability	yes	yes	yes	yes	yes	no	yes	yes	yes
Ability to show the position of a detectable mutation	yes	yes	no	yes	yes	yes	yes	with difficulty	yes
Ability to tower background with DNA polymerase and DNA ligase recycling reaction	no	no	no	no	with difficulty	no	no	no	yes
Advantage to multiplex samples of same color	unknown	no	with difficulty	yes	unknown	no	no	no	yes
Advantage to analyze targets of 1 Kbp–3 Kbp	unknown	unknown	with difficulty	yes, up to 1 Kbp	unknown	no	no	no	yes

REFERENCES

1. Modrich, P. (1994) *Science* 266, 1959–1960.
2. Su, S.-S., Lahue, R. S., Au, K. G., and Moldrich, P. (1988) *J. Biol. Chem.* 263, 5057–5061.
3. Finkelstein, E., Afek U., Gross, E., Aharoni, N., Rosenberg, L., and Halevy, S. (1994) *International Journal of Dermatology* 33, 116–118.
4. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Analytical Chemistry* 150, 76–85.
5. Gregory D. J., Culp, D. J., and Jahnke, M. R. (1990) *Analytical Biochem.* 185, 324–330.
6. Yeung, A. T., Mattes, W. B., Oh, E. Y., and Grossman, L. (1983) *Proc. Nat. Acad Sci. USA,* 80, 6157–6161.
7. Yeung, A. T., Dinehart, W. J. and Jones, B. K. (1988) *Nucleic Acids Res,* 16, 4539–4554.
8. Lyamichev, V., Brow, M. A. D., and Dahlberg, J. E. (1993) *Science* 260, 778–783.
9. Ramotar, D., Auchincloss, A. H., and Fraser, M. J. (1987) *J. Biol. Chem.* 262, 425–31.
10. Chow, T. Y.-K., and Resnick, M. A. (1987) *J. Biol. Chem.* 262, 17659–17667.
11. Wyen, N. V., Erdei, S., and Farkas, G. L. (1971) *Biochem Biophys. Acta.* 232, 472–83.
12. Brown, P. H., and Ho, D. T. (1987) *Eur. J. Biochem.* 168, 357–364.
13. Hanson, D. M. and Fairley, J. L. (1969) *J. Biol. Chem.* 244, 2440–2449.
14. Nucleases, eds. Linn, S. M., Lloyd, R. S., and Roberts, R. J. Cold Spring Harbor Laboratory Press, 1993.
15. Holloman, W. K., Rowe, T. C., and Rusche, J. R (1981) *J. Biol. Chem.* 256, 5087–5094.
16. Badman, R. (1972) *Genetic Res., Camb.* 20, 213–229.
17. Shank T. E., Rhodes, C. Rigby, P. W. J., and Berg, P. (1975) *Proc. Nat. Acad. Sci. USA,* 72, 989–993.
18. Maekawa, K., Tsunasawa, S., Dibo, G., and Saklyama, F. (1991) *Eur. J. Biochem.* 200, 651–661.
19. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* 15, 4457–4462.
20. Kroeker, W. D., Kowalski, D., and Laskowski, M. Sr. (1976) *Biochemistry* 15, 4463–4467.
21. Ardelt, W., and Laskowski, M., Sr. (1971) *Biochem. Biophys. Res., Commun.* 44, 1205–1211.
22. Kowalski, D. (1984) *Nucleic Acids. Res.* 12, 7071–7086.
23. Strickland, J. A., Marzilli L. G., Puckett, Jr., J. M., and Doetsch, P. W. (1991) *Biochemistry* 30, 9749–9756.
24. Doetsch, P. W., McCray, W. H., Lee, K., Bettler, D. R., and Valenzuela, M. R. L. (1988) *Nucleic Acids Res.* 16, 6935–6952.
25. Caren, P. R., Kushner, S. R., and Grossman, L, (1985) *Proc. Nat. Acad. Sci. USA* 82, 4925–4929.
26. Yeh, Y.-C., Liu, H.-F., Ellis, C. A., and Lu, A.-L. (1994) *J. Biol. Chem.* 269, 15498–15504.
27. Welch, W. J., Gerrels, J. I., Thomas, G. P., Lin, J. J.-L., and Feramisco, J. R (1983) *J. Biol. Chem.* 258, 7102–7111.
28. Jackson, S. P., and Tjian, R (1989) *Proc. Nat. Acad. Sci. USA* 86, 1781–1785.
29. Jackson, S. P., and Tjian, R (1988) *Cell* 55, 125–133.
30. He, J., and Furmanski, P. (1995) *Nature* 373, 721–724.
31. Peeples, M. E. (1988) *Virology* 162, 255–259.
32. Buckley, A., and Gould, E. A. (1988) *J. Gen. Virology,* 69, 1913–1920.
33. Gauffre, A., Viron, A., Barel, M., Hermann, J., Puvion, E., and Frade, R. (1992) *Molecular Immunology* 29, 1113–1120.
34. Jones, B. K. and Yeung, A. T. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8410–8414.
35. Sarker, et al., (1992) *Nucleic Acids Research* 20:871–878.
36. Meyers, R. M. et al., (1986) CSHSQB 52:275.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 13

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC 60

CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA 120

TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA 180

TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC 240

ACATATTTTG CAAATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT 300

GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC 360

AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT 420

ATGCAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG 480

AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG 540

AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGGAA 600

CTGTGAGAAC TCTGAGGACA AAGCAGCGGA TACAACCTCA AAAGACGTCT GTCTACATTG 660

AATTGGGATC TGATTCTTCT GAAGATACCG TTAATAAGGC AACTTATTGC AGTGTGGGAG 720

ATCAAGAATT GTTACAAATC ACCCCTCAAG GAACCAGGGA TGAAATCAGT TTGGATTCTG 780

CAAAAAAGGC TGCTTGTGAA TTTTCTGAGA CGGATGTAAC AAATACTGAA CATCATCAAC 840

CCAGTAATAA TGATTTGAAC ACCACTGAGA AGCGTGCAGC TGAGAGGCAT CCAGAAAAGT 900

ATCAGGGTAG TTCTGTTTCA AACTTGCATG TGGAGCCATG TGGCACAAAT ACTCATGCCA 960

GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTCACTAA AGACAGAATG AATGTAGAAA 1020

AGGCTGAATT CTGTAATAAA AGCAAACAGC CTGGCTTAGC AAGGAGCCAA CATAACAGAT 1080

GGGCTGGAAG TAAGGAAACA TGTAATGATA GGCGGACTCC CAGCACAGAA AAAAAGGTAG 1140

ATCTGAATGC TGATCCCCTG TGTGAGAGAA AAGAATGGAA TAAGCAGAAA CTGCCATGCT 1200

CAGAGAATCC TAGAGATACT GAAGATGTTC CTTGGATAAC ACTAAATAGC AGCATTCAGA 1260

AAGTTAATGA GTGGTTTTCC AGAAGTGATG AACTGTTAGG TTCTGATGAC TCACATGATG 1320

GGGAGTCTGA ATCAAATGCC AAAGTAGCTG ATGTATTGGA CGTTCTAAAT GAGGTAGATG 1380

AATATTCTGG TTCTTCAGAG AAAATAGACT TACTGGCCAG TGATCCTCAT GAGGCTTTAA 1440

TATGTAAAAG TGAAAGAGTT CACTCCAAAT CAGTAGAGAG TAATATTGAA GACAAAATAT 1500

TTGGGAAAAC CTATCGGAAG AAGGCAAGCC TCCCCAACTT AAGCCATGTA ACTGAAAATC 1560

TAATTATAGG AGCATTTGTT ACTGAGCCAC AGATAATACA AGAGCGTCCC CTCACAAATA 1620

AATTAAAGCG TAAAAGGAGA CCTACATCAG GCCTTCATCC TGAGGATTTT ATCAAGAAAG 1680

-continued

CAGATTTGGC AGTTCAAAAG ACTCCTGAAA TGATAAATCA GGGAACTAAC CAAACGGAGC	1740
AGAATGGTCA AGTGATGAAT ATTACTAATA GTGGTCATGA GAATAAAACA AAAGGTGATT	1800
CTATTCAGAA TGAGAAAAAT CCTAACCCAA TAGAATCACT CGAAAAAGAA TCTGCTTTCA	1860
AAACGAAAGC TGAACCTATA AGCAGCAGTA TAAGCAATAT GGAACTCGAA TTAAATATCC	1920
ACAATTCAAA AGCACCTAAA AAGAATAGGC TGAGGAGGAA GTCTTCTACC AGGCATATTC	1980
ATGCGCTTGA ACTAGTAGTC AGTAGAAATC TAAGCCCACC TAATTGTACT GAATTGCAAA	2040
TTGATAGTTG TTCTAGCAGT GAAGAGATAA AGAAAAAAAA GTACAACCAA ATGCCAGTCA	2100
GGCACAGCAG AAACCTACAA CTCATGGAAG GTAAAGAACC TGCAACTGGA GCCAAGAAGA	2160
GTAACAAGCC AAATGAACAG ACAAGTAAAA GACATGACAG CGATACTTTC CCAGAGCTGA	2220
AGTTAACAAA TGCACCTGGT TCTTTTACTA AGTGTTCAAA TACCAGTGAA CTTAAAGAAT	2280
TTGTCAATCC TAGCCTTCCA AGAGAAGAAA AAGAAGAGAA ACTAGAAACA GTTAAAGTGT	2340
CTAATAATGC TGAAGACCCC AAAGATCTCA TGTTAAGTGG AGAAAGGGTT TTGCAAACTG	2400
AAAGATCTGT AGAGAGTAGC AGTATTTCAT TGGTACCTGG TACTGATTAT GGCACTCAGG	2460
AAAGTATCTC GTTACTGGAA GTTAGCACTC TAGGGAAGGC AAAAACAGAA CCAAATAAAT	2520
GTGTGAGTCA GTGTGCAGCA TTTGAAAACC CCAAGGGACT AATTCATGGT TGTTCCAAAG	2580
ATAATAGAAA TGACACAGAA GGCTTTAAGT ATCCATTGGG ACATGAAGTT AACCACAGTC	2640
GGGAAACAAG CATAGAAATG GAAGAAAGTG AACTTGATGC TCAGTATTTG CAGAATACAT	2700
TCAAGGTTTC AAAGCGCCAG TCATTTGCTC CGTTTTCAAA TCCAGGAAAT GCAGAAGAGG	2760
AATGTGCAAC ATTCTCTGCC CACTCTGGGT CCTTAAAGAA ACAAAGTCCA AAAGTCACTT	2820
TTGAATGTGA ACAAAAGGAA GAAAATCAAG GAAAGAATGA GTCTAATATC AAGCCTGTAC	2880
AGACAGTTAA TATCACTGCA GGCTTTCCTG TGGTTGGTCA GAAAGATAAG CCAGTTGATA	2940
ATGCCAAATG TAGTATCAAA GGAGGCTCTA GGTTTTGTCT ATCATCTCAG TTCAGAGGCA	3000
ACGAAACTGG ACTCATTACT CCAAATAAAC ATGGACTTTT ACAAAACCCA TATCGTATAC	3060
CACCACTTTT TCCCATCAAG TCATTTGTTA AAACTAAATG TAAGAAAAAT CTGCTAGAGG	3120
AAAACTTTGA GGAACATTCA ATGTCACCTG AAAGAGAAAT GGGAAATGAG AACATTCCAA	3180
GTACAGTGAG CACAATTAGC CGTAATAACA TTAGAGAAAA TGTTTTTAAA GAAGCCAGCT	3240
CAAGCAATAT TAATGAAGTA GGTTCCAGTA CTAATGAAGT GGGCTCCAGT ATTAATGAAA	3300
TAGGTTCCAG TGATGAAAAC ATTCAAGCAG AACTAGGTAG AAACAGAGGG CCAAAATTGA	3360
ATGCTATGCT TAGATTAGGG GTTTTGCAAC CTGAGGTCTA TAAACAAAGT CTTCCTGGAA	3420
GTAATTGTAA GCATCCTGAA ATAAAAAAGC AAGAATATGA AGAAGTAGTT CAGACTGTTA	3480
ATACAGATTT CTCTCCATAT CTGATTTCAG ATAACTTAGA ACAGCCTATG GGAAGTAGTC	3540
ATGCATCTCA GGTTTGTTCT GAGACACCTG ATGACCTGTT AGATGATGGT GAAATAAAGG	3600
AAGATACTAG TTTTGCTGAA AATGACATTA AGGAAAGTTC TGCTGTTTTT AGCAAAAGCG	3660
TCCAGAAAGG AGAGCTTAGC AGGAGTCCTA GCCCTTTCAC CCATACACAT TTGGCTCAGG	3720
GTTACCGAAG AGGGGCCAAG AAATTAGAGT CCTCAGAAGA GAACTTATCT AGTGAGGATG	3780
AAGAGCTTCC CTGCTTCCAA CACTTGTTAT TTGGTAAAGT AAACAATATA CCTTCTCAGT	3840
CTACTAGGCA TAGCACCGTT GCTACCGAGT GTCTGTCTAA GAACACAGAG GAGAATTTAT	3900
TATCATTGAA GAATAGCTTA AATGACTGCA GTAACCAGGT AATATTGGCA AAGGCATCTC	3960
AGGAACATCA CCTTAGTGAG GAAACAAAAT GTTCTGCTAG CTTGTTTTCT TCACAGTGCA	4020
GTGAATTGGA AGACTTGACT GCAAATACAA ACACCCAGGA TCCTTTCTTG ATTGGTTCTT	4080

-continued

CCAAACAAAT GAGGCATCAG TCTGAAAGCC AGGGAGTTGG TCTGAGTGAC AAGGAATTGG	4140
TTTCAGATGA TGAAGAAAGA GGAACGGGCT TGGAAGAAAA TAATCAAGAA GAGCAAAGCA	4200
TGGATTCAAA CTTAGGTGAA GCAGCATCTG GGTGTGAGAG TGAAACAAGC GTCTCTGAAG	4260
ACTGCTCAGG GCTATCCTCT CAGAGTGACA TTTTAACCAC TCAGCAGAGG GATACCATGC	4320
AACATAACCT GATAAAGCTC CAGCAGGAAA TGGCTGAACT AGAAGCTGTG TTAGAACAGC	4380
ATGGGAGCCA GCCTTCTAAC AGCTACCCTT CCATCATAAG TGACTCTTCT GCCCTTGAGG	4440
ACCTGCGAAA TCCAGAACAA AGCACATCAG AAAAAGCAGT ATTAACTTCA CAGAAAAGTA	4500
GTGAATACCC TATAAGCCAG AATCCAGAAG GCCTTTCTGC TGACAAGTTT GAGGTGTCTG	4560
CAGATAGTTC TACCAGTAAA AATAAAGAAC CAGGAGTGGA AAGGTCATCC CCTTCTAAAT	4620
GCCCATCATT AGATGATAGG TGGTACATGC ACAGTTGCTC TGGGAGTCTT CAGAATAGAA	4680
ACTACCCATC TCAAGAGGAG CTCATTAAGG TTGTTGATGT GGAGGAGCAA CAGCTGGAAG	4740
AGTCTGGGCC ACACGATTTG ACGGAAACAT CTTACTTGCC AAGGCAAGAT CTAGAGGGAA	4800
CCCCTTACCT GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCCTGAATCT GATCCTTCTG	4860
AAGACAGAGC CCCAGAGTCA GCTCGTGTTG GCAACATACC ATCTTCAACC TCTGCATTGA	4920
AAGTTCCCCA ATTGAAAGTT GCAGAATCTG CCCAGAGTCC AGCTGCTGCT CATACTACTG	4980
ATACTGCTGG GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG	5040
CTTCAACAGA AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG	5100
AATTTATGCT CGTGTACAAG TTTGCCAGAA AACACCACAT CACTTTAACT AATCTAATTA	5160
CTGAAGAGAC TACTCATGTT GTTATGAAAA CAGATGCTGA GTTTGTGTGT GAACGGACAC	5220
TGAAATATTT TCTAGGAATT GCGGGAGGAA AATGGGTAGT TAGCTATTTC TGGGTGACCC	5280
AGTCTATTAA AGAAAGAAAA ATGCTGAATG AGCATGATTT TGAAGTCAGA GGAGATGTGG	5340
TCAATGGAAG AAACCACCAA GGTCCAAAGC GAGCAAGAGA ATCCCAGGAC AGAAAGATCT	5400
TCAGGGGGCT AGAAATCTGT TGCTATGGGC CCTTCACCAA CATGCCCACA GATCAACTGG	5460
AATGGATGGT ACAGCTGTGT GGTGCTTCTG TGGTGAAGGA GCTTTCATCA TTCACCCTTG	5520
GCACAGGTGT CCACCCAATT GTGGTTGTGC AGCCAGATGC CTGGACAGAG GACAATGGCT	5580
TCCATGCAAT TGGGCAGATG TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA	5640
GTGTAGCACT CTACCAGTGC CAGGAGCTGG ACACCTACCT GATACCCCAG ATCCCCCACA	5700
GCCACTACTG A	5711

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of substrate Nos. 1, 12, 13, and 14."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(i x) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..63
(D) OTHER INFORMATION: /product="Substrate No. 1" / standard_name= "top strand 5'to 3'"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

-continued

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGCGTCAT GTGTGGAATT GTGATTAAAA 60

TCG 63

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Bottom strand of Substrate Nos. 1, 2, 3, 4, 5, 7, 10, 15"

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCGATTTTAA TCACAATTCC ACACATGACG CGAGCCGGAA GCATAAAGTG AACTAGCATG 60

ACG 63

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate No. 2"

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGGCGTCA TGTGTGGAAT TGTGATTAAA 60

ATCG 64

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate No. 3"

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGTCGTCA TGTGTGGAAT TGTGATTAAA 60

ATCG 64

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid

-continued

(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate No. 4."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGACGTCA TGTGTGGAAT TGTGATTAAA 60

ATCG 64

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate No. 5."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCGCCGTCA TGTGTGGAAT TGTGATTAAA 60

ATCG 64

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate Nos. 6, 7, 8, 18."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCACGTCAT GTGTGGAATT GTGATTAAAA 60

TCG 63

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate Nos. 9, 10, 11, 19."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCCCGTCAT GTGTGGAATT GTGATTAAAA 60

TCG 63

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Top strand of Substrate Nos. 15, 16, 17, 20."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCGTCATGCT AGTTCACTTT ATGCTTCCGG CTCTCGTCAT GTGTGGAATT GTGATTAAAA 60

TCG 63

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Bottom strand of Substrate Nos. 6, 9, 12, 20."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCGATTTTAA TCACAATTCC ACACATCACG AGAGCCGGAA GCATAAAGTG AACTAGCATG 60

ACG 63

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Bottom strand of Substrate Nos. 8, 13, 16, 19."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCGATTTTAA TCACAATTCC ACACATCACG GGAGCCGGAA GCATAAAGTG AACTAGCATG 60

ACG 63

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Bottom strand of Substrate Nos. 11, 14, 17, 18."

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCGATTTTAA TCACAATTCC ACACATCACG TGAGCCGGAA GCATAAAGTG AACTAGCATG 60

ACG 63

What is claimed is:

1. A method for determining a mutation in a target sequence of a single stranded polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including said target sequence, wherein said polynucleotides are amplified, labeled with a detectable marker, hybridized to one another, subjected to the activity of an endonuclease and analyzed for the presence of said mutation, the improvement comprising the use of a mismatch endonuclease enzyme of plant origin, the activity of said enzyme comprising:

a) detection of all mismatches whether known or unknown between said hybridized polynucleotides, said detection occurring over a pH range of 5–9, said enzyme exhibiting substantial activity over the entire pH range;

b) catalytic formation of a substantially single-stranded nick at a target sequence containing a mismatch; and

c) recognition of a mutation in a target polynucleotide sequence, said recognition being substantially unaffected by flanking polynucleotide sequences.

2. The method as claimed in claim 1 wherein said endonuclease is from celery.

3. The method as claimed in claim 1 wherein said polynucleotide is DNA.

4. The method as claimed in claim 2 wherein the sequences subjected to said endonuclease activity are further subjected to the activity of a protein, said protein being selected from the group consisting of DNA ligase, DNA polymerase, DNA helicase, 3'-5' DNA Exonuclease, DNA binding proteins that bind to DNA termini or a combination of said proteins, thereby reducing non-specific DNA cleavage.

5. The method as claimed in claim 2 wherein the sequences subjected to said endonuclease activity are further subjected to DNA polymerase activity, so as to reduce non-specific DNA cleavage.

6. The method as claimed in claim 2 wherein target DNA is analyzed in the presence of a multiplicity of pooled samples.

7. The method as claimed in claim 2 wherein said polynucleotide is cDNA.

8. The method as claimed in claim 1, wherein said polynucleotides are analyzed on a DNA sequencing gel thereby identifying the location of the mutation in a target DNA strand relative to DNA sequencing molecular weight markers.

9. The method as claimed in claim 1 wherein said determination is employed as an assay for detection of cancer.

10. The method as claimed in claim 1 wherein said determination is employed as an assay for detection of birth defects.

11. A method for determining a mutation in a target sequence of single stranded polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including said target sequence, wherein said polynucleotides are amplified, labeled with a detectable marker, hybridized to one another, exposed to endonuclease and analyzed for the presence of said mutation, the improvement comprising the use of a mismatch endonuclease enzyme from celery, the activity of said enzyme comprising:

a) detection of all mismatches whether known or unknown between said hybridized polynucleotides, said detection occurring over a pH range of 5–9, said enzyme exhibiting substantial activity over the entire pH range;

b) catalytic formation of a substantially single-stranded nick at a target sequence containing a mismatch;

c) recognition of a mutation in a target polynucleotide said recognition being substantially unaffected by flanking polynucleotide sequences; and

d) recognition of polynucleotide loops and insertions between said hybridized polynucleotides.

12. The method as claimed in claim 2 wherein the sequences subjected to said endonuclease activity are further subjected to the activity of a protein, said protein being selected from the group consisting of DNA ligase, DNA polymerase, DNA helicase, 3'-5' DNA Exonuclease, DNA binding proteins that bind to DNA termini or a combination of said proteins, thereby stimulating turnover of said endonuclease.

13. The method as claimed in claim 2 wherein said sequences subjected to said endonuclease activity are further subjected to DNA polymerase activity, thereby stimulating turnover of said endonuclease.

14. A mismatch endonuclease enzyme for determining a mutation in a target sequence of single stranded mammalian polynucleotide with reference to a non-mutated sequence in a polynucleotide that is hybridizable with the polynucleotide including said target sequence, said enzyme being isolated from a plant source and effective to:

a) detect all mismatches, whether known or unknown between said hybridized polynucleotides, said detection occurring over a pH range of 5–9, said enzyme exhibiting substantial activity over the entire pH range;

b) recognize polynucleotide loops and insertions in said hybridized polynucleotides;

c) catalyze formation of a substantially single-stranded nick at the DNA site containing a mismatch;

d) recognize a mutation in a target polynucleotide sequence, said recognition being substantially unaffected by flanking DNA sequences.

15. An enzyme as claimed in claim **14**, wherein said enzyme is CEL I.

16. An enzyme as claimed in claim **14**, said enzyme being in substantially pure form.

17. An enzyme as claimed in claim **15**, said enzyme being in substantially pure form.

* * * * *

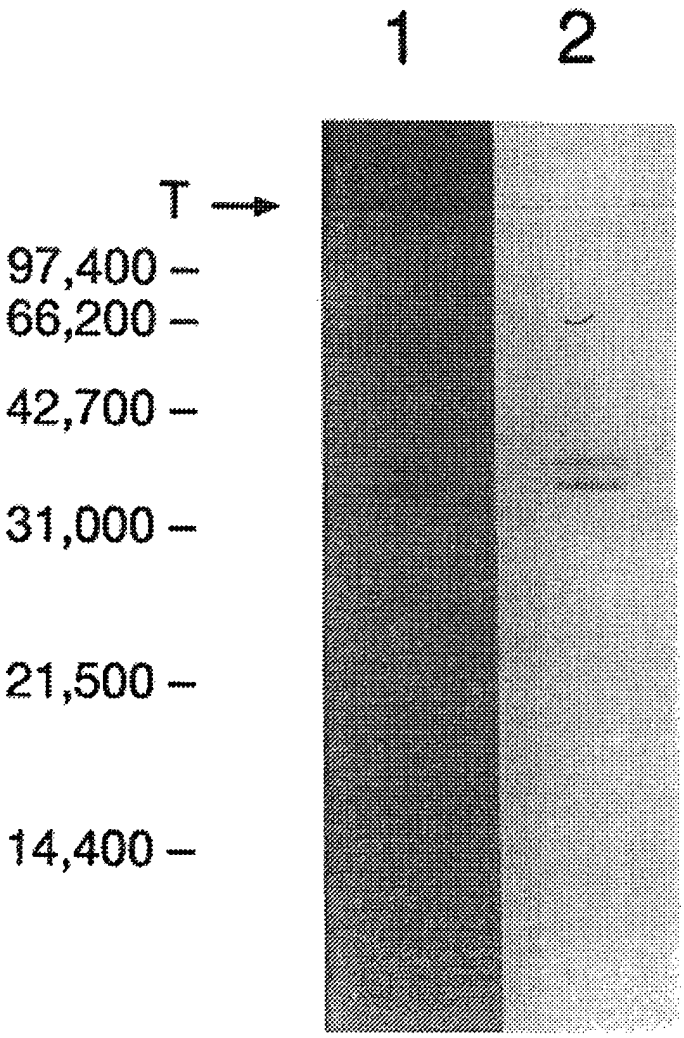
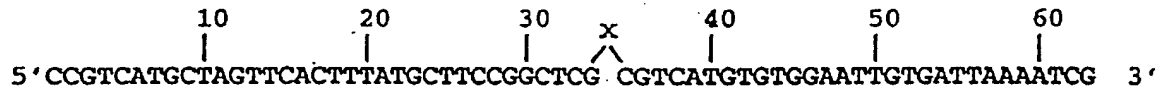


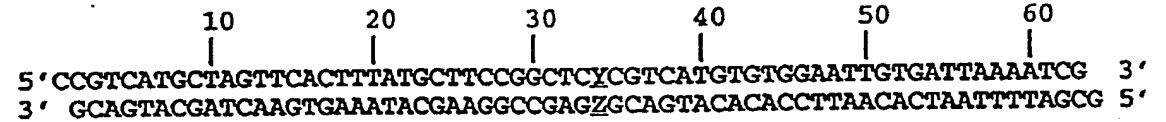
FIG. 1

Fig. 2A



Substrate #	1	2	3	4	5
Loop X	-	G	T	A	C

Fig. 2B



Substrate #	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	A	A	A	C	C	C	G	G	G	T	T	T	A	C	T
Z	A	C	G	A	C	T	A	G	T	C	G	T	T	G	A

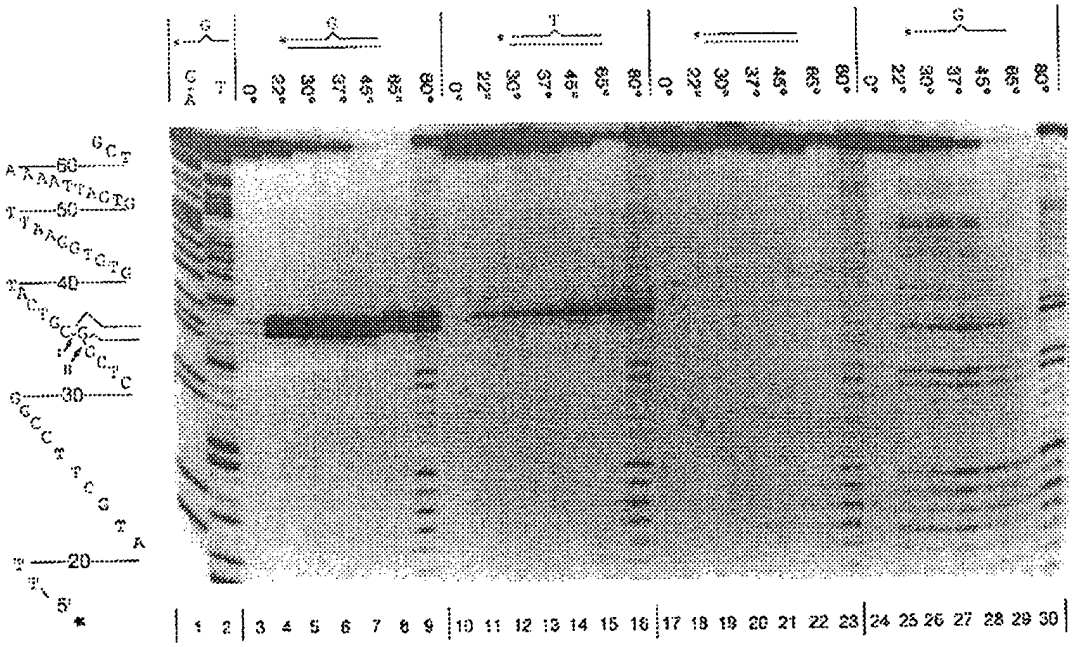

G
G
T
G
C/A
T
0°
22°
30°
37°
45°
65°
80°
50
40
30
20
5'
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

FIG. 3

FIG. 4A

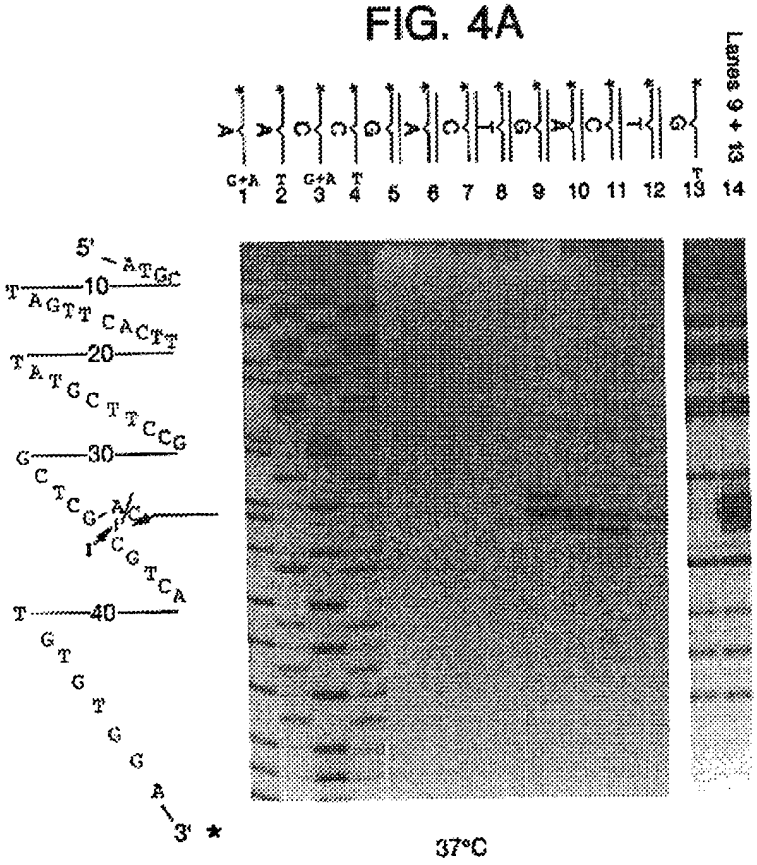

Lanes 9 + 13
G+A 1
T 2
G+A 3
T 4
5
6
7
8
9
10
11
12
T 13
14
5'
10
20
30
40
3'
-1
37°C

FIG. 4B

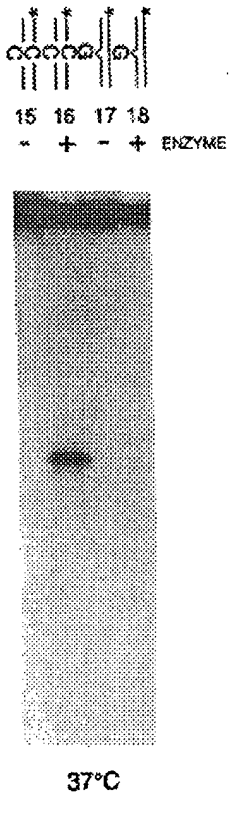

15
16
17
18
- + - + ENZYME
37°C

FIG. 5A

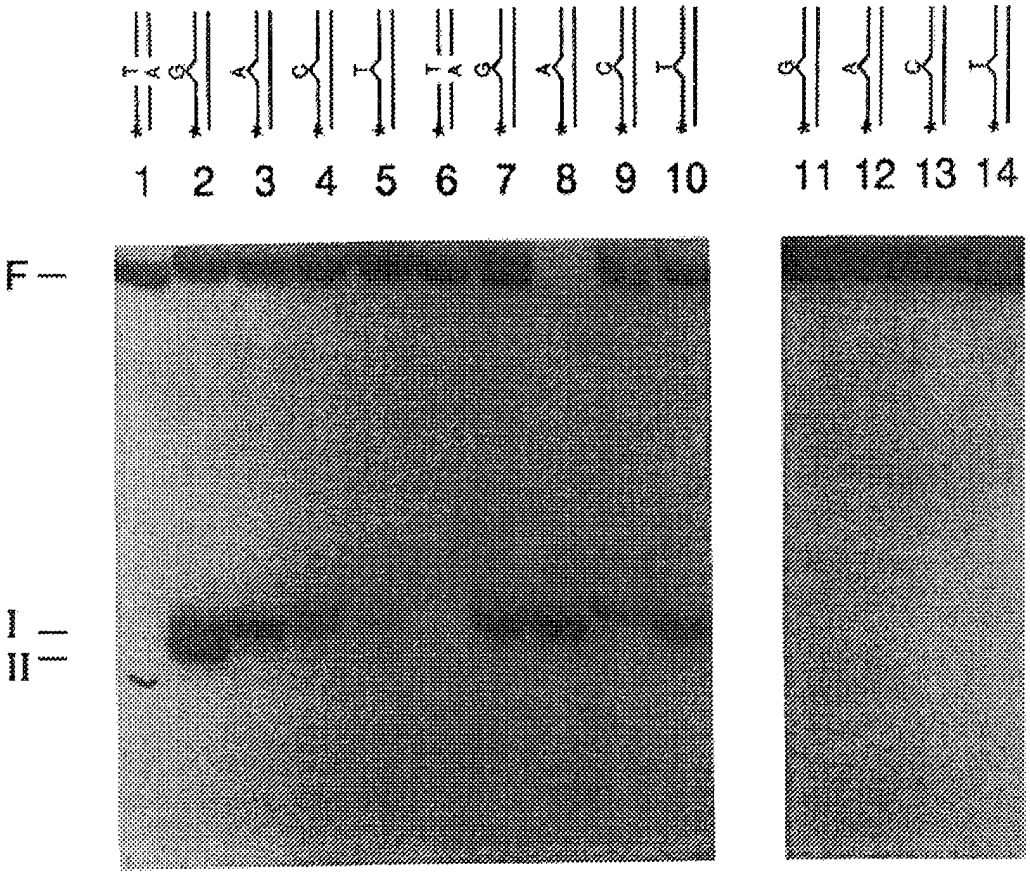


FIG. 5B

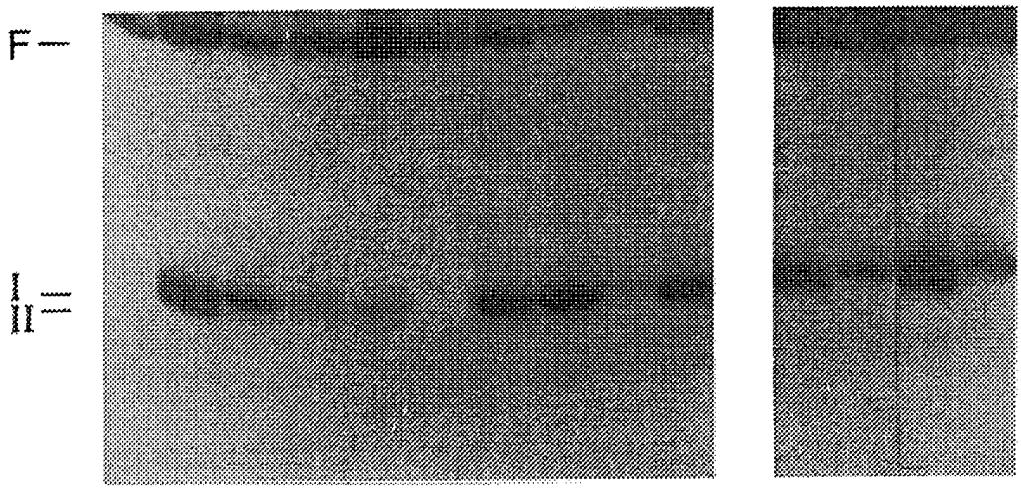


FIG. 5C

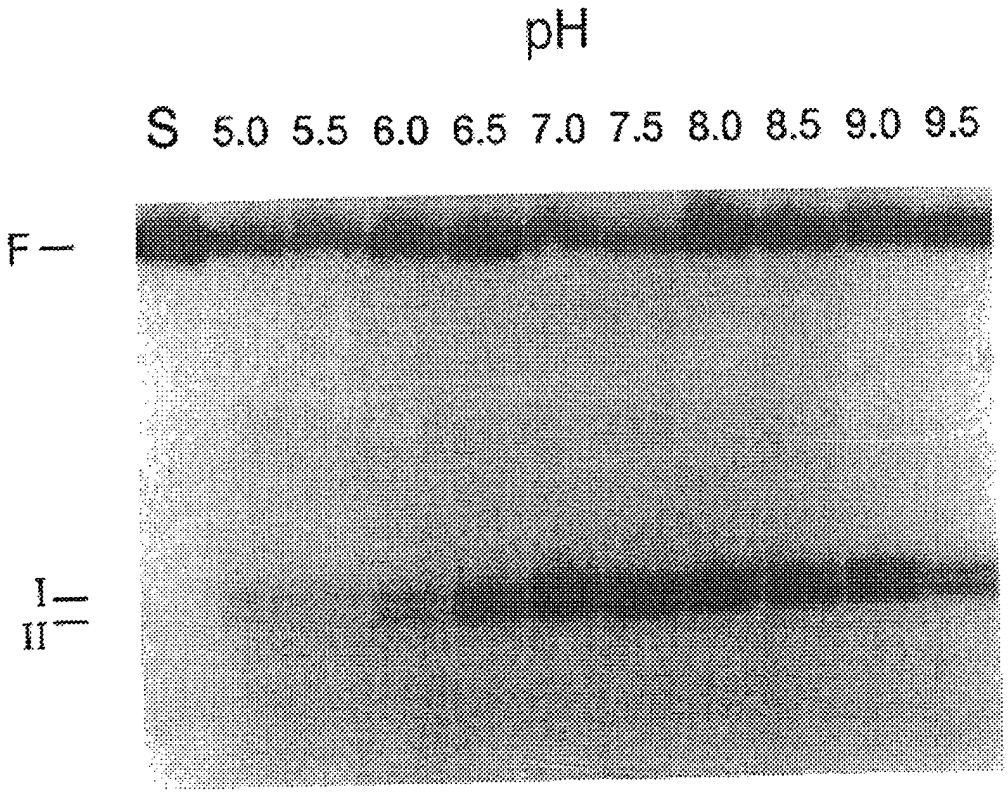

pH
S 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5
F
I
II

FIG. 6A

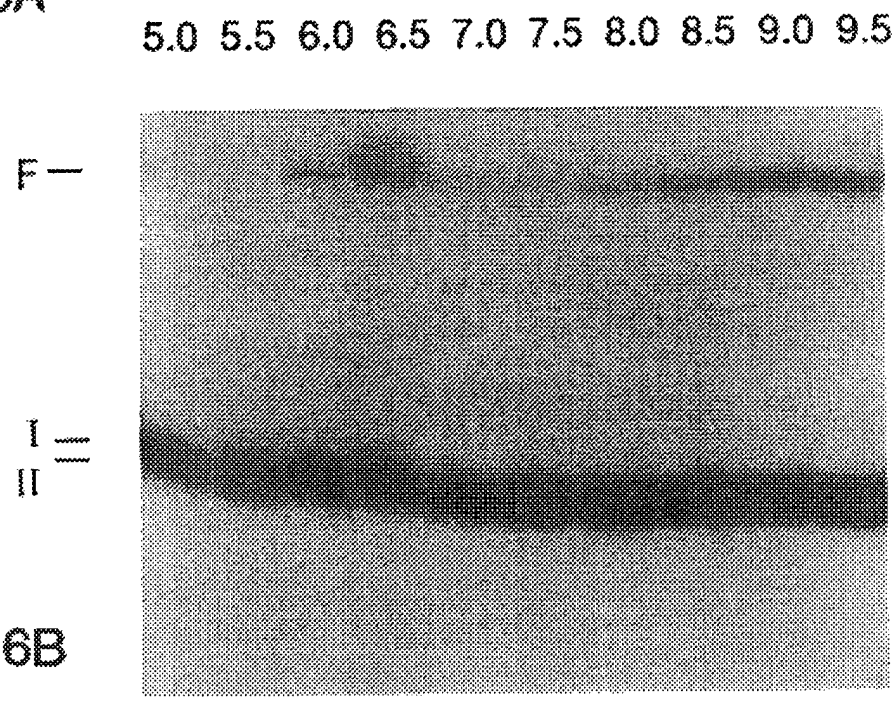

5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5
F
I
II

FIG. 6B

FIG. 7A

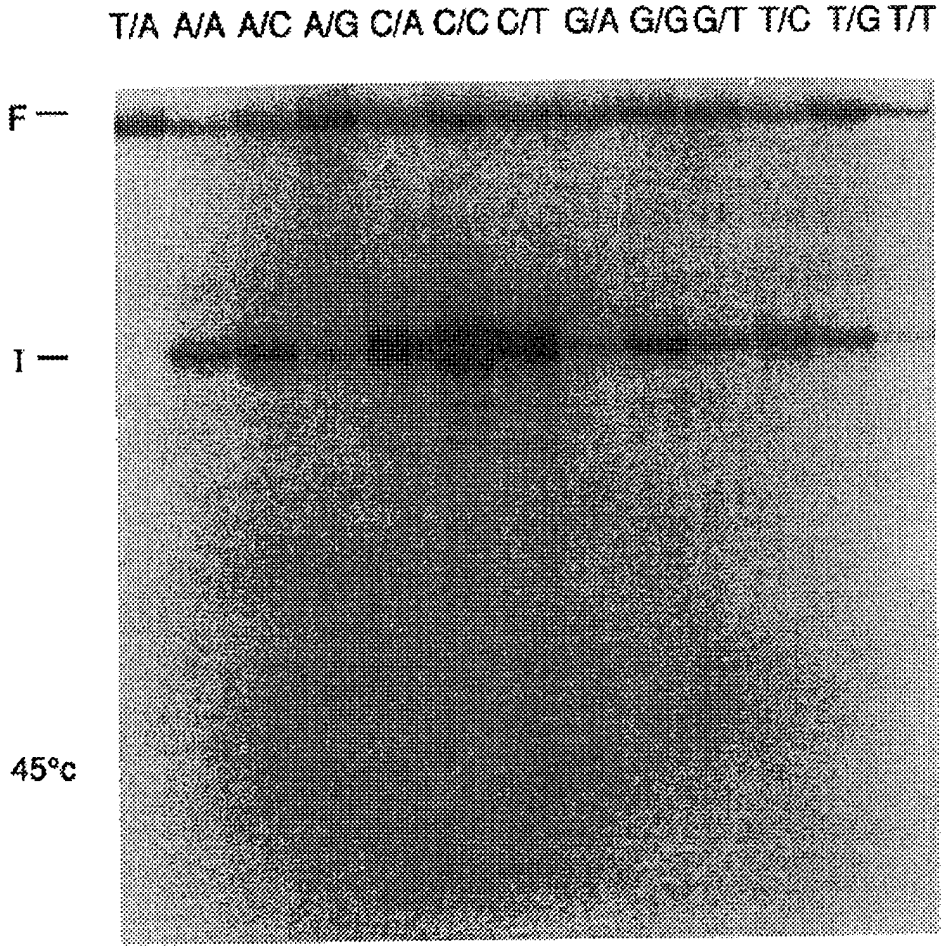
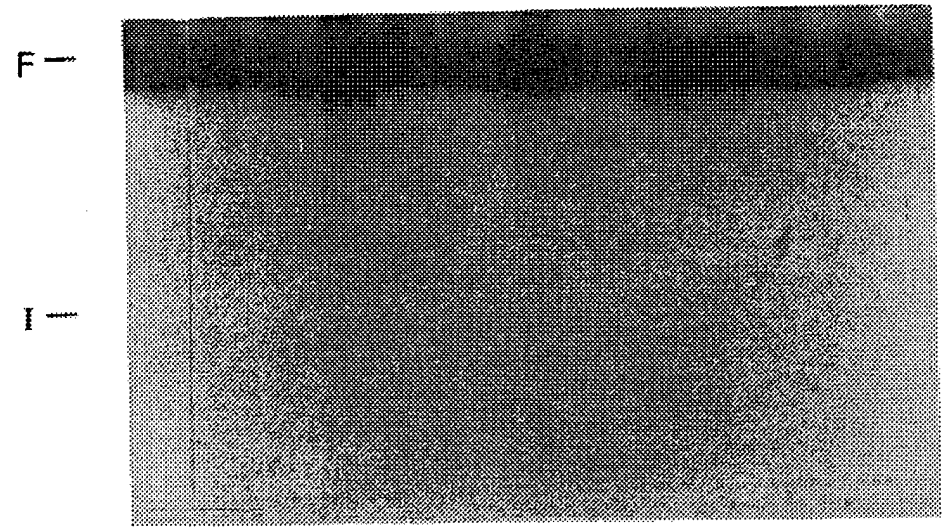


FIG. 7B



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

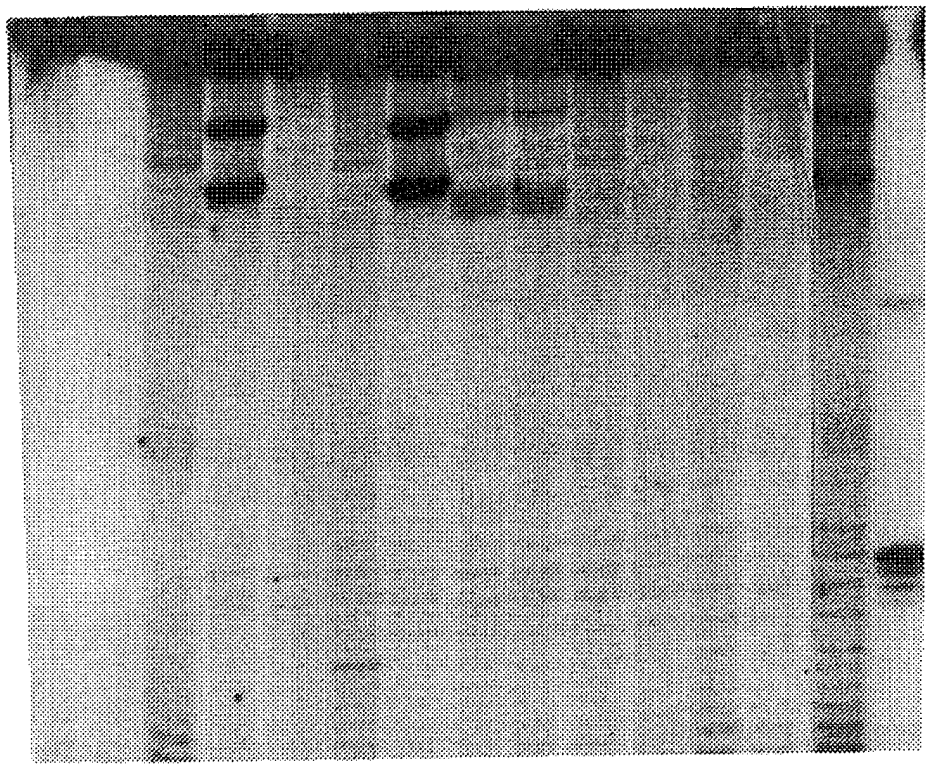


FIG. 8

1 2 3 4 5 6 7 8 9 10 11

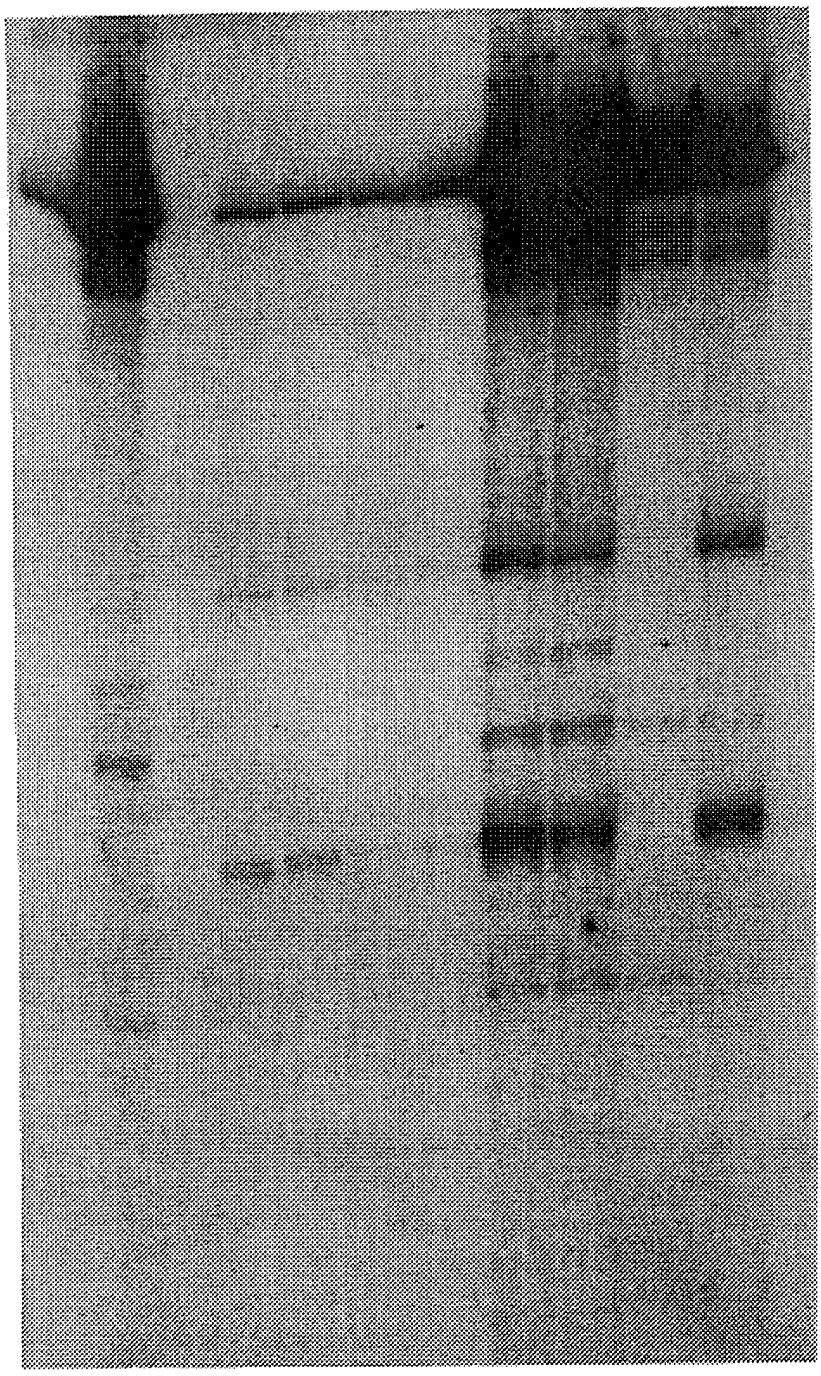


FIG. 9

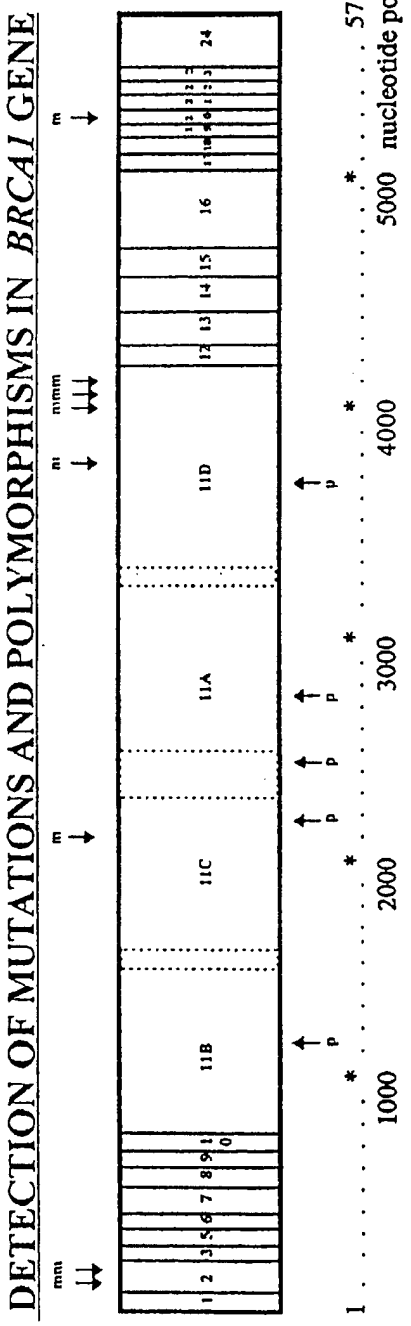

DETECTION OF MUTATIONS AND POLYMORPHISMS IN BRCA1 GENE
mmi
m
mi mmm
m
1 2 3 5 6 7 8 9 10
11B
11C
11A
11D
12 13 14 15 16 24
p p p p p
1 . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * 5711
1000 2000 3000 4000 5000 nucleotide position
= overlapping regions of Exon 11 sections p = polymorphism m = mutation

Figure 10

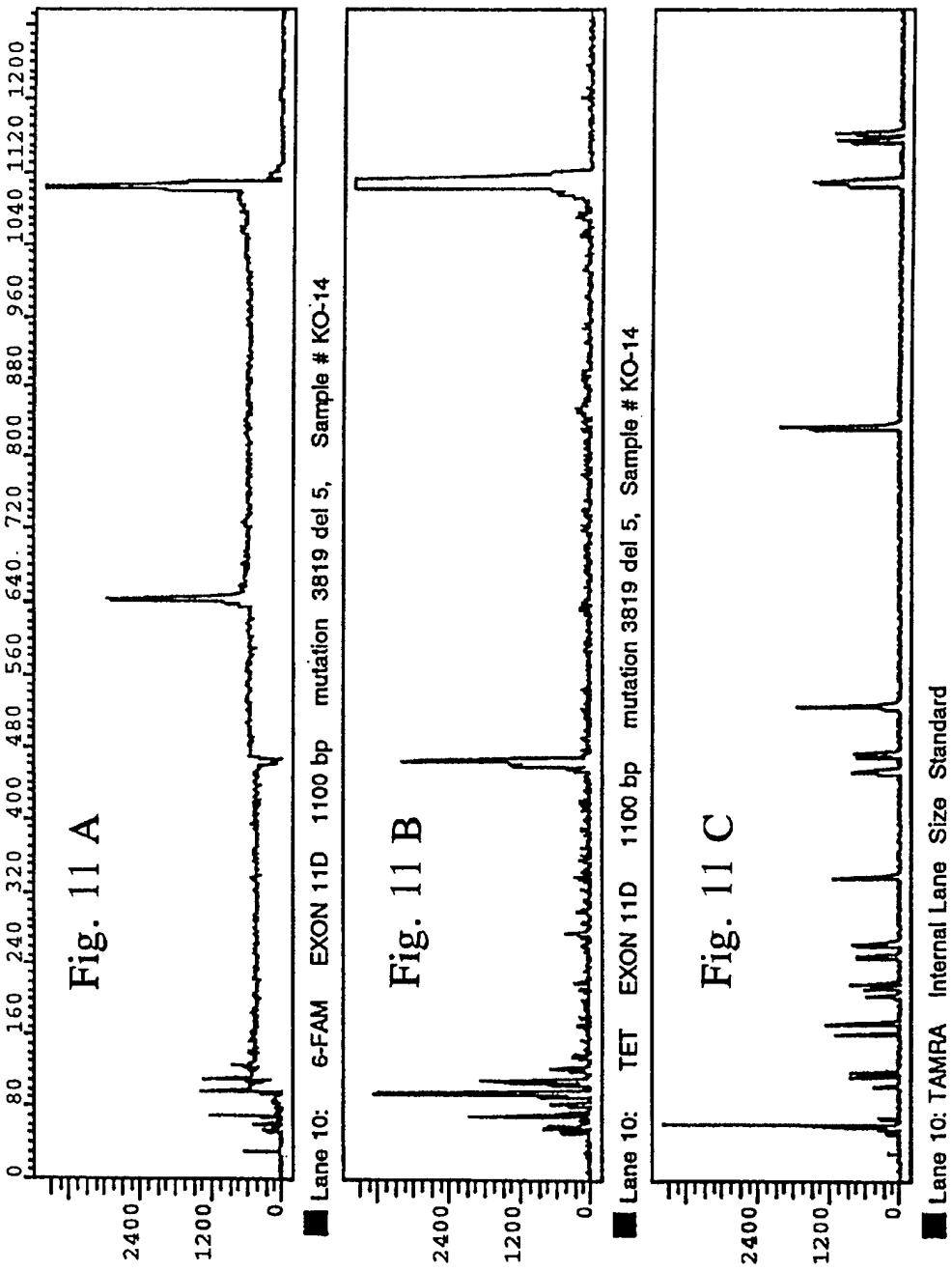


Fig. 11 A

Fig. 11 B

Fig. 11 C

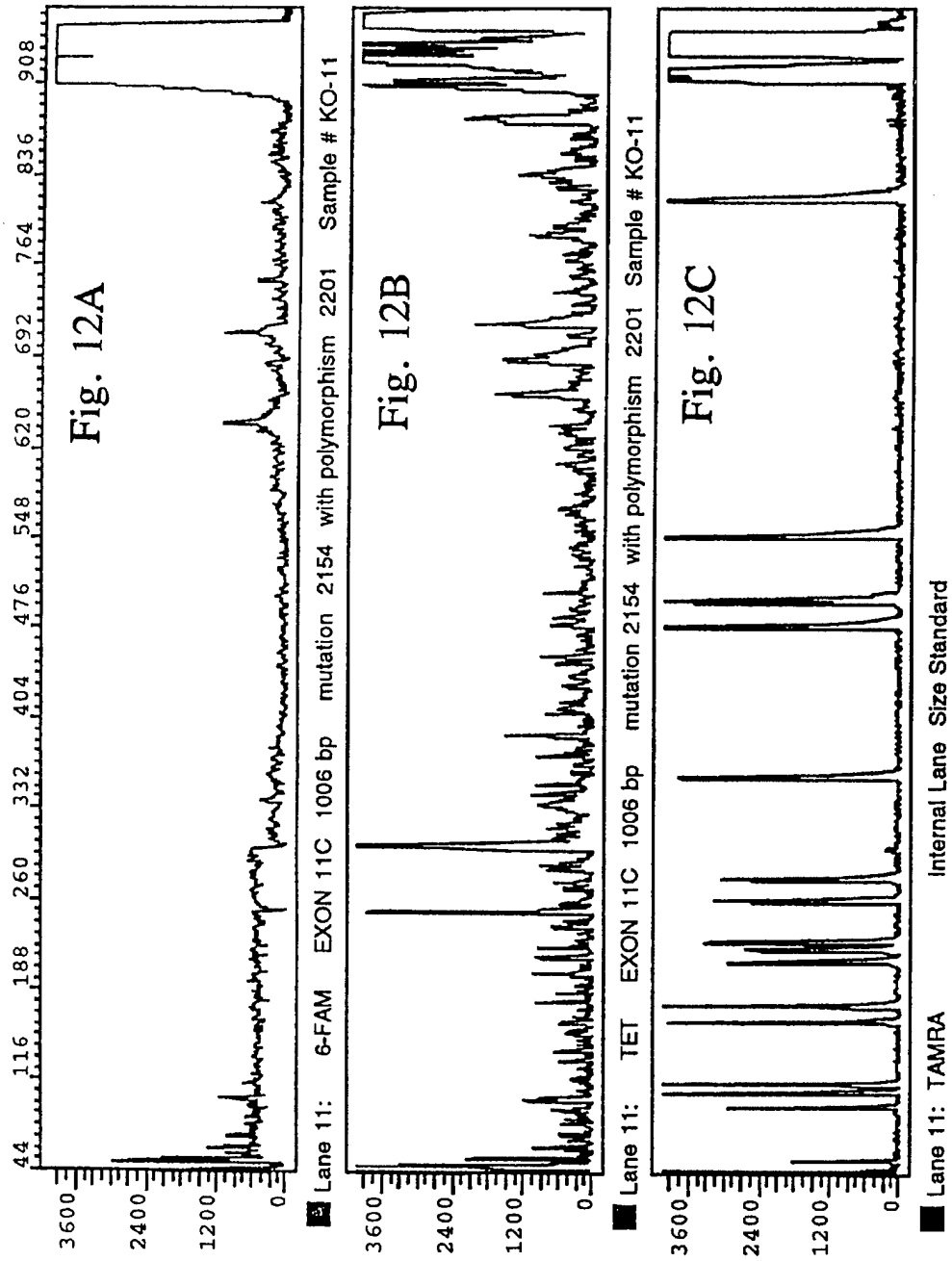


Fig. 12A

Fig. 12B

Fig. 12C

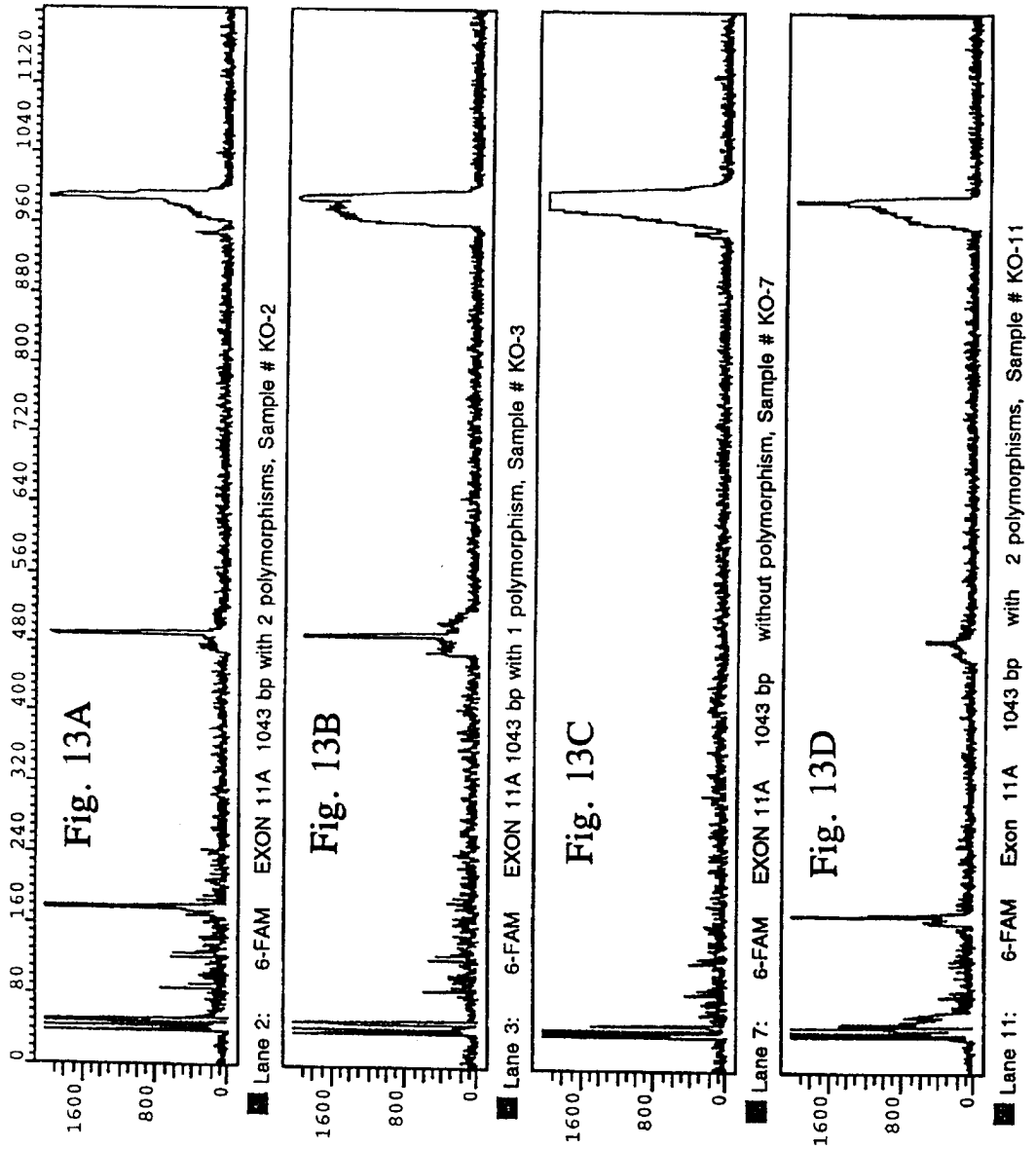


Fig. 13A
Lane 2: 6-FAM EXON 11A 1043 bp with 2 polymorphisms, Sample # KO-2
Fig. 13B
Lane 3: 6-FAM EXON 11A 1043 bp with 1 polymorphism, Sample # KO-3
Fig. 13C
Lane 7: 6-FAM EXON 11A 1043 bp without polymorphism, Sample # KO-7
Fig. 13D
Lane 11: 6-FAM Exon 11A 1043 bp with 2 polymorphisms, Sample # KO-11
0 80 160 240 320 400 480 560 640 720 800 880 960 1040 1120
1600
800
0